(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 October 2002 (03.10.2002)

PCT

(10) International Publication Number
WO 02/077267 A2

(51) International Patent Classification[7]: C12Q

(21) International Application Number: PCT/US02/09611

(22) International Filing Date: 27 March 2002 (27.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|---|---|---|
| 60/279,228 | 27 March 2001 (27.03.2001) | US |
| 60/291,501 | 15 May 2001 (15.05.2001) | US |

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

| | |
|---|---|
| US | 60/279,228 (CIP) |
| Filed on | 27 March 2001 (27.03.2001) |
| US | 60/291,501 (CIP) |
| Filed on | 15 May 2001 (15.05.2001) |

(71) Applicant *(for all designated States except US)*: **CORVAS INTERNATIONAL, INC.** [US/US]; 3030 Science Park Road, San Diego, CA 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MADISON, Edwin, L.** [US/US]; 11005 Cedarcrest Way, San Diego, CA 92121 (US). **ONG, Edgar, O.** [CA/US]; 10738 Glendover Lane, San Diego, CA 92126 (US).

(74) Agents: **SEIDMAN, Stephanie, L.** et al.; Heller Ehrman White & McAuliffe LLP, 4350 La Jolla Village Drive, San Diego, CA 92122-1246 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17:**

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

**Published:**

— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRAN SERINE PROTEASE 9, THE ENCODED POLYPEPTIDES AND METHODS BASED THEREON

(57) Abstract: Provided herein are type II transmembrane serine protease 9 (MTSP9) polypeptides. Zymogen and activated forms of these polypeptides as well as single and two chain forms of the protease domaine are also provided. Methods using the polypeptides to identify compounds that modulate the protease activity of an MTSP9 are provided.

WO 02/077267 A2

# NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9, THE ENCODED POLYPEPTIDES AND METHODS BASED THEREON

## RELATED APPLICATIONS

Benefit of priority is claimed to U.S. provisional application Serial No. 60/279,228, filed March 27, 2001, Edwin L. Madison and Edgar O. Ong, entitled "NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 9, THE ENCODED PROTEINS AND METHODS BASED THEREON", and to U.S. provisional application Serial No. 60/291,501, filed May 15, 2001, to Edwin L. Madison and Edgar O. Ong, entitled "NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 9, THE ENCODED PROTEINS AND METHODS BASED THEREON." Where permitted, the subject matter of each of these application is incorporated in its entirety by reference thereto.

## FIELD OF INVENTION

Nucleic acid molecules that encode proteases and portions thereof, particularly protease domains are provided. Also provided are prognostic, diagnostic and therapeutic methods using the proteases and domains thereof and the encoding nucleic acid molecules.

## BACKGROUND OF THE INVENTION AND OBJECTS THEREOF

Cancer, which is a leading cause of death in the United States, is characterized by an increase in the number of abnormal neoplastic cells, which proliferate to form a tumor mass, the invasion of adjacent tissues by these neoplastic tumor cells, and the generation of malignant cells that metastasize via the blood or lymphatic system to regional lymph nodes and to distant sites. Among the hallmarks of cancer is a breakdown in the communication among tumor cells and their environment. Normal cells do not divide in the absence of stimulatory signals and cease dividing in the presence of inhibitory signals. Growth-stimulatory and growth-inhibitory signals, are routinely exchanged between cells within a tissue. In a cancerous, or neoplastic, state, a cell acquires the ability to "override" these signals and to proliferate under conditions in which normal cells do not grow.

**Type II Transmembrane Serine Proteases (TTSPs)**

In addition to the MMPs, serine proteases have been implicated in neoplastic disease progression. Most serine proteases, which are either secreted enzymes or are sequestered in cytoplasmic storage organelles, have roles in blood coagulation, wound healing, digestion, immune responses and tumor invasion and metastasis. A class of cell surface proteins designated type II transmembrane serine proteases, which are membrane-anchored proteins with additional extracellular domains, has been identified. As cell surface proteins, they are positioned to play a role in intracellular signal transduction and in mediating cell surface proteolytic events.

Cell surface proteolysis is a mechanism for the generation of biologically active proteins that mediate a variety of cellular functions. Membrane-associated proteases include membrane-type metalloproteinases (MT-MMP), ADAMs (proteases that contain disintegrin-like and metalloproteinase domains) and the TTSPs. In mammals, at least 17 members of the TTSP family are known, including seven in humans (see, Hooper *et al.* (2001) *J. Biol. Chem. 276*:857-860). These include: corin (accession nos. AF133845 and AB013874; see, Yan *et al.* (1999) *J. Biol. Chem. 274*:14926-14938; Tomia *et al.* (1998) *J. Biochem. 124*:784-789; Uan *et al.* (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*:8525-8529); enterpeptidase (also designated enterokinase; accession no. U09860 for the human protein; see, Kitamoto *et al.* (1995) *Biochem. 27*: 4562-4568; Yahagi *et al.* (1996) *Biochem. Biophys. Res. Commun. 219*:806-812; Kitamoto *et al.* (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*:7588-7592; Matsushima *et al.* (1994) *J. Biol. Chem. 269*:19976-19982;); human airway trypsin-like protease (HAT; accession no. AB002134; see Yamaoka *et al. J. Biol. Chem. 273*:11894-11901); MTSP1 and matriptase (also called TADG-15; see SEQ ID Nos. 1 and 2; accession nos. AF133086/AF118224, AF04280022; Takeuchi *et al.* (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*:11054-1161; Lin *et al.* (1999) *J. Biol. Chem. 274:*18231-18236; Takeuchi *et al.* (2000) *J. Biol. Chem. 275*:26333-26342; and Kim *et al.* (1999) *Immunogenetics 49*:420-429); hepsin (see, accession nos. M18930, AF030065, X70900; Leytus *et al.* (1988) *Biochem. 27*: 11895-11901; Vu *et al.* (1997) *J. Biol. Chem. 272*:31315-31320;

Prostate-specific antigen (PSA), a kallikrein-like serine protease, degrades extracellular matrix glycoproteins fibronectin and laminin, and, has been postulated to facilitate invasion by prostate cancer cells (Webber *et al.* (1995) *Clin. Cancer Res.*, 1(10):1089-94). Blocking PSA proteolytic activity with PSA-specific monoclonal antibodies results in a dose-dependent decrease *in vitro* in the invasion of the reconstituted basement membrane Matrigel by LNCaP human prostate carcinoma cells which secrete high levels of PSA.

Hepsin, a cell surface serine protease identified in hepatoma cells, is overexpressed in ovarian cancer (Tanimoto *et al.* (1997) *Cancer Res.*, 57(14):2884-7). The hepsin transcript appears to be abundant in carcinoma tissue and is almost never expressed in normal adult tissue, including normal ovary. It has been suggested that hepsin is frequently overexpressed in ovarian tumors and therefore can be a candidate protease in the invasive process and growth capacity of ovarian tumor cells.

A serine protease-like gene, designated normal epithelial cell-specific 1 (NES1) (Liu et al., *Cancer Res.*, 56(14):3371-9 (1996)) has been identified. Although expression of the NES1 mRNA is observed in all normal and immortalized nontumorigenic epithelial cell lines, the majority of human breast cancer cell lines show a drastic reduction or a complete lack of its expression. The structural similarity of NES1 to polypeptides known to regulate growth factor activity and a negative correlation of NES1 expression with breast oncogenesis suggest a direct or indirect role for this protease-like gene product in the suppression of tumorigenesis.

Hence transmembrane serine proteases appear to be involved in the etiology and pathogenesis of tumors. There is a need to further elucidate their role in these processes and to identify additional transmembrane proteases. Therefore, it is an object herein to provide transmembrane serine protease (MTSP) proteins and nucleic acids encoding such MTSP proteases that are involved in the regulation of or participate in tumorigenesis and/or carcinogenesis. It is also an object herein to provide prognostic, diagnostic and therapeutic screening methods using such proteases and the nucleic acids encoding such proteases.

and in SEQ ID No. 16. The protein sequence and encoding nucleic acid sequence of the full-length MTSP9 are set forth in SEQ ID Nos. 18 and 17.

Also provided are nucleic acid molecules that hybridize to such MTSP9-encoding nucleic acid along their full-length or along at least about 70%, 80% or 90% of the full-length and encode the protease domain or portion thereof are provided. Hybridization is generally effected under conditions of at least low, generally at least moderate, and often high stringency.

The isolated nucleic acid fragment is DNA, including genomic or cDNA, or is RNA, or can include other components, such as protein nucleic acid or other nucleotide analogs. The isolated nucleic acid may include additional components, such as heterologous or native promoters, and other transcriptional and translational regulatory sequences, these genes may be linked to other genes, such as reporter genes or other indicator genes or genes that encode indicators.

Also provided is an isolated nucleic acid molecule that includes the sequence of molecules that is complementary to the nucleotide sequence encoding MTSP9 or the portion thereof.

Also provided are fragments thereof or oligonucleotides that can be used as probes or primers and that contain at least about 10, 14, 16 nucleotides, generally less than 1000 or less than or equal to 100, set forth in SEQ ID No. 5 or 17 (or the complement thereof); or contain at least about 30 nucleotides (or the complement thereof) or contain oligonucleotides that hybridize along their full-length (or at least about 70, 80 or 90% thereof) to any such fragments or oligonucleotides. The length of the fragments are a function of the purpose for which they are used and/or the complexity of the genome of interest. Generally probes and primers contain less than about 50, 150 or 500 nucleotides.

Also provided are plasmids containing any of the nucleic acid molecules provided herein. Cells containing the plasmids are also provided. Such cells include, but are not limited to, bacterial cells, yeast cells, fungal cells, plant cells, insect cells and animal cells.

Also provided is a method of producing MTSP9 by growing the above-described cells under conditions whereby the MTSP9 is expressed by the cells,

In a specific embodiment, a nucleic acid that encodes a MTSP, designated MTSP9 is provided. In particular, the nucleic acid includes the sequence of nucleotides set forth in SEQ ID No. 5, particularly set forth as nucleotides 31-729 of SEQ ID No. 5, or SEQ ID No. 17 or a portion there of that encodes a catalytically active polypeptide.

Also provided are nucleic acid molecules that hybridize under conditions of at least low stringency, generally moderate stringency, more typically high stringency to the SEQ ID No. 5 or 17 or degenerates thereof.

In one embodiment, the isolated nucleic acid fragment hybridizes to a nucleic acid molecule containing the nucleotide sequence set forth in SEQ ID No: 5 or 17 (or degenerates thereof) under high stringency conditions, in one embodiments contains the sequence of nucleotides set forth in SEQ ID Nos. 5 and 17. A full-length MTSP9 is set forth in SEQ ID No. 18 and is encoded by SEQ ID No. 17 or degenerates thereof.

Also provided are muteins of the single chain protease domain of MTSP9 particularly muteins in which the Cys residue in the protease domain that is free (*i.e.*, does not form disulfide linkages with any other Cys residue in the protease domain) is substituted with another amino acid substitution, typically, although not necessarily, with a conservative amino acid substitution or a substitution that does not eliminate the activity, and muteins in which a glycosylation site(s) is eliminated.

Hence muteins in which one or more of the Cys residues, particularly, a residue that is paired in the activated two form, but unpaired in the protease domain alone (*i.e.*, the Cys a residue position 26 (see SEQ ID Nos. 5, 6 and 16) in the protease domain), is/are replaced with any amino acid, typically, although not necessarily, a conservative amino acid residue, such as Ser, are contemplated. Muteins of MTSP9, particularly those in which Cys residues, such as the unpaired Cys in the single chain protease domain, is replaced with another amino acid that does not eliminate the activity, are provided. Muteins in which other

for breast, prostate and colon cancer. The expression and/or activation of MTSP9 on or in the vicinity of a cell or in a bodily fluid in a subject can be a marker for breast, prostate, lung, colon and other cancers.

In certain embodiments, the MTSP9 polypeptide is detectable in a body fluid at a level that differs from its level in body fluids in a subject not having a tumor. In other embodiments, the polypeptide is present in a tumor; and a substrate or cofactor for the polypeptide is expressed at levels that differ from its level of expression in a non-tumor cell in the same type of tissue. In other embodiments, the level of expression and/or activity of the MTSP9 polypeptide in tumor cells differs from its level of expression and/or activity in non-tumor cells. In other embodiments, the MTSP9 is present in a tumor; and a substrate or cofactor for the MTSP9 is expressed at levels that differ from its level of expression in a non-tumor cell in the same type of tissue.

Also provided are methods for screening for compounds that modulate the activity of MTSP9. The compounds are identified by contacting them with the MTSP9 or protease domain thereof and a substrate for the MTSP9. A change in the amount of substrate cleaved in the presence of the compounds compared to that in the absence of the compound indicates that the compound modulates the activity of the MTSP9. Such compounds are selected for further analyses or for use to modulate the activity of the MTSP9, such as inhibitors or agonists. The compounds can also be identified by contacting the substrates with a cell that expresses the MTSP9 or the extracellular domain or proteolytically active portion thereof.

Also provided herein are methods of modulating the activity of the MTSP9 and screening for compounds that modulate, including inhibit, antagonize, agonize or otherwise alter the activity of the MTSP9. Of particular interest is the extracellular domain of MTSP9 that includes the proteolytic (catalytic) portion of the protein.

Cells, combinations, kits and articles of manufacture containing the MTSP9 polypeptides, domains thereof, or encoding nucleic acids are also provided herein. Methods of expressing the encoded MTSP9 polypeptide and portions thereof using the cells are also provided, as are cells that express

(i) inhibits activation of the single-chain zymogen form of MTSP9;
(ii) inhibits activity of the two-chain or single-chain form; and
(iii) inhibits dimerization of the protein.

The forms can be full length or truncated forms, including but not limited to, the protease domain resulting from cleavage at the activation cleavage site (between amino acids $R_{185}$ and $I_{186}$); or from expression of the protease domain or catalytically active portions thereof.

Pharmaceutical composition containing the protease domain and/or full-length or other domain of an MTSP9 polypeptide are provided herein in a pharmaceutically acceptable carrier or excipient are provided herein.

Also provided are articles of manufacture that contain MTSP9 polypeptide and protease domains of MTSP9 in single chain forms or activated forms. The articles contain a) packaging material; b) the polypeptide (or encoding nucleic acid), particularly the single chain protease domain thereof; and c) a label indicating that the article is for using ins assays for identifying modulators of the activities of an MTSP9 polypeptide is provided herein.

Conjugates containing a) an MTSP9 polypeptide or protease domain in a single or two chain form; and b) a targeting agent linked to the MTSP directly or via a linker, wherein the agent facilitates: i) affinity isolation or purification of the conjugate; ii) attachment of the conjugate to a surface; iii) detection of the conjugate; or iv) targeted delivery to a selected tissue or cell, is provided herein. The conjugate can contain a plurality of agents linked thereto. The conjugate can be a chemical conjugate; and it can be a fusion protein. The targeting agent can be a protein or peptide fragment. The protein or peptide fragment can include a protein binding sequence, a nucleic acid binding sequence, a lipid binding sequence, a polysaccharide binding sequence, or a metal binding sequence.

Combinations are provided herein. The combination can include: a) an inhibitor of the activity of an MTSP9; and b) an anti-cancer treatment or agent. The MTSP inhibitor and the anti-cancer agent can be formulated in a single pharmaceutical composition or each is formulated in a separate pharmaceutical composition. The MTSP9 inhibitor can be an antibody or a fragment or binding

Recombinant and transgenic animals can be produced by homologous recombination and non-homologous recombination methods.

Methods of gene therapy are provided. Such methods can be effected administering *in vivo* or *ex vivo* an inactivating form of the MTSP9 or by administering an MTSP-encoding nucleic acid molecule are also provided.

Also provided are methods of treatments of tumors by administering a prodrug that is activated by MTSP9 that is expressed or active in tumor cells, particularly those in which its functional activity in tumor cells is greater than in non-tumor cells. The prodrug is administered and, upon administration, active MTSP9 expressed on cells cleaves the prodrug and releases active drug in the vicinity of the tumor cells. The active anti-cancer drug accumulates in the vicinity of the tumor. This is particularly useful in instances in which MTSP9 is expressed or active in greater quantity, higher level or predominantly in tumor cells compared to other cells.

Also provided are methods of diagnosing the presence of a pre-malignant lesion, a malignancy, or other pathologic condition in a subject, by obtaining a biological sample from the subject; exposing it to a detectable agent that binds to a two-chain and/or single-chain form of MTSP9, where the pathological condition is characterized by the presence or absence of the two-chain and/or single-chain form.

Methods of inhibiting tumor invasion or metastasis or treating a malignant or pre-malignant condition by administering an agent that inhibits activation of the zymogen form of MTSP9 or an activity of the activated form are provided. The conditions include, but are not limited to, a condition, such as a tumor, of the breast, cervix, prostate, lung, ovary or colon.

## DETAILED DESCRIPTION

### A. DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which the invention(s) belong. All patents, patent applications, published applications and publications, Genbank sequences, websites and other published materials referred to throughout the entire disclosure herein, unless noted

exemplary MTSPs and/or domains thereof are set forth, for example in U.S. application Serial No. 09/776,191 (SEQ ID Nos. 1-12, 49, 50 and 61-72 therein, published as International PCT application No. WO 01/57194). The term also encompass MTSPs with amino acid substitutions that do not substantially alter activity of each member and also encompasses splice variants thereof. Suitable substitutions, including, although not necessarily, conservative substitutions of amino acids, are known to those of skill in this art and can be made without eliminating the biological activity, such as the catalytic activity, of the resulting molecule.

As used herein an MTSP9, whenever referenced herein, includes at least one or all of or any combination of:

a polypeptide encoded by the sequence of nucleotides set forth in SEQ ID No. 17 or by a sequence of nucleotides that includes nucleotides that encode amino acids acids 11-232 of SEQ ID No. 6;

a polypeptide encoded by a sequence of nucleotides that hybridizes under conditions of low, moderate or high stringency to the sequence of nucleotides set forth in is set forth as nucleotides 31-729 SEQ ID No. 5 or as SEQ ID No. 17;

a polypeptide that includes the sequence of amino acids set forth as amino acids 11-232 of SEQ ID No. 6;

a polypeptide that includes a sequence of amino acids having at least about 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity with the sequence of amino acids set forth in SEQ ID No. 17 or 18 or as amino acids 11-232 of SEQ ID No. 6; and/or

a polypeptide encoded by a splice variant of the MTSP9 set forth in SEQ ID No. 17.

In particular, the MTSP9 polypeptide, with the protease domains as indicated in SEQ ID No. 5, 6, 16,17 and 18, is provided. The polypeptide is a single or two chain polypeptide. Smaller portions thereof that retain protease activity are also provided. The protease domains from MTSPs vary in size and constitution, including insertions and deletions in surface loops. They retain

a polypeptide encoded by a sequence of nucleotides that hybridizes under conditions of low, moderate or high stringency to the sequence of nucleotides set forth in SEQ ID No. 15 or 17;

a polypeptide that includes the sequence of amino acids set forth in SEQ ID No. 6, 16 or 18;

a polypeptide that includes a sequence of amino acids having at least about 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity with the sequence of amino acids set forth in SEQ ID No. 6, 16; or 18; and/or

a protease domain of a polypeptide encoded by a splice variant of the MTSP9.

The protease domains of MTSPs vary in size and constitution, including insertions and deletions in surface loops. They retain conserved structure, including at least one of the active site triad, primary specificity pocket, oxyanion hole and/or other features of serine protease domains of proteases. Thus, for purposes herein, the protease domain is a portion of an MTSP, as defined herein, and is homologous to a domain of other MTSP. As with the larger class of enzymes of the chymotrypsin (S1) fold (see, *e.g.*, Internet accessible MEROPS data base), the MTSP protease domains share a high degree of amino acid sequence identity. The His, Asp and Ser residues necessary for activity are present in conserved motifs. The activation site, whose cleavage creates the N-terminus of the protease domain in the two-chain forms is located in a conserved motif and readily can be identified.

By active form is meant a form active *in vivo* and/or *in vitro*. As described herein, the protease domain also can exist as a two-chain form. It is shown herein that, at least *in vitro*, the single chain forms of the SPs and the catalytic domains or proteolytically active portions thereof (typically C-terminal truncations) exhibit protease activity. Hence provided herein are isolated single chain forms of the protease domains of SPs and their use in *in vitro* drug screening assays for identification of agents that modulate the activity thereof.

As used herein, catalytic activity refers to the activity of the SP as a serine protease. Function of the SP refers to its function in tumor biology, including promotion of or involvement in initiation, growth or progression of tumors, and also roles in signal transduction. Catalytic activity refers to the activity of the SP as a protease as assessed in *in vitro* proteolytic assays that detect proteolysis of a selected substrate.

As used herein, a zymogen is an inactive precursor of a proteolytic enzyme. Such precursors are generally larger, although not necessarily larger than the active form. With reference to serine proteases, zymogens are converted to active enzymes by specific cleavage, including catalytic and autocatalytic cleavage, or by binding of an activating co-factor, which generates the an active enzyme. A zymogen, thus, is an enzymatically inactive protein that is converted to a proteolytic enzyme by the action of an activator.

As used herein, "disease or disorder" refers to a pathological condition in an organism resulting from, *e.g.*, infection or genetic defect, and characterized by identifiable symptoms.

As used herein, neoplasm (neoplasia) refers to abnormal new growth, and thus means the same as *tumor*, which can be benign or malignant. Unlike *hyperplasia*, neoplastic proliferation persists even in the absence of the original stimulus.

As used herein, neoplastic disease refers to any disorder involving cancer, including tumor development, growth, metastasis and progression.

As used herein, cancer refers to a general term for diseases caused by any type of malignant tumor.

As used herein, malignant, as applies to tumors, refers to primary tumors that have the capacity of *metastasis* with loss of *growth control* and *positional control*.

As used herein, an anti-cancer agent (used interchangeable with "anti-tumor or anti-neoplastic agent") refers to any agents used in the anti-cancer treatment. These include any agents, when used alone or in combination with other compounds, that can alleviate, reduce, ameliorate, prevent, or place or maintain in a state of remission of clinical symptoms or diagnostic markers

outweighs the influence of angiogenesis inhibitors. As used herein, deficient angiogenesis refers to pathological angiogenesis associated with disorders where there is a defect in normal angiogenesis resulting in aberrant angiogenesis or an absence or substantial reduction in angiogenesis.

As used herein, the protease domain of an SP protein refers to the protease domain of an SP that exhibits proteolytic activity. Hence it is at least the minimal portion of the protein that exhibits proteolytic activity as assessed by standard assays *in vitro*. It refers, herein, to a single chain form and also the two chain activated form (where the two chain form is intended it will be so-noted). Exemplary protease domains include at least a sufficient portion of sequences of amino acids set forth in SEQ ID No. 6 (encoded by nucleotides in SEQ ID No. 5) to exhibit protease activity.

Also contemplated are nucleic acid molecules that encode a polypeptide that has proteolytic activity in an *in vitro* proteolysis assay and that have at least 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity with the full-length of a protease domain of an MTSP9 polypeptide, or that hybridize along their full-length or along at least about 70%, 80% or 90% of the full-length to a nucleic acids that encode a protease domain, particularly under conditions of moderate, generally high, stringency.

For the protease domains, residues at the N-terminus can be critical for activity. It is shown herein that the protease domain of the single chain form of the MTSP9 protease is catalytically active. Hence the protease domain generally requires the N-terminal amino acids thereof for activity; the C-terminus portion can be truncated. The amount that can be removed can be determined empirically by testing the polypeptide for protease activity in an *in vitro* assay that assesses catalytic cleavage.

Hence smaller portions of the protease domains, particularly the single chain domains, thereof that retain protease activity are contemplated. Such smaller versions generally are C-terminal truncated versions of the protease domains. The protease domains vary in size and constitution, including insertions and deletions in surface loops. Such domains exhibit conserved

program package (Devereux, J., *et al., Nucleic Acids Research 12(I)*:387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., *et al., J Molec Biol 215*:403 (1990); Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo *et al.* (1988) *SIAM J Applied Math 48*:1073). For example, the BLAST function of the National Center for Biotechnology Information database can be used to determine identity. Other commercially or publicly available programs include, DNAStar "MegAlign" program (Madison, WI) and the University of Wisconsin Genetics Computer Group (UWG) "Gap" program (Madison WI)). Percent homology or identity of proteins and/or nucleic acid molecules can be determined, for example, by comparing sequence information using a GAP computer program (*e.g.*, Needleman *et al.* (1970) *J. Mol. Biol.* 48:443, as revised by Smith and Waterman ((1981) *Adv. Appl. Math.* 2:482). Briefly, the GAP program defines similarity as the number of aligned symbols (i.e., nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. Default parameters for the GAP program can include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) and the weighted comparison matrix of Gribskov *et al.* (1986) *Nucl. Acids Res.* 14:6745, as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3) no penalty for end gaps. Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide.

As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of 100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons can be made between a test and reference polynucleotides. Such differences can be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they can be clustered in one or more

normally produced in the mammalian host or that is not produced in therapeutically effective amounts or at a therapeutically useful time. The heterologous nucleic acid, such as DNA, encoding the therapeutic product can be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof. Genetic therapy can also involve delivery of an inhibitor or repressor or other modulator of gene expression.

As used herein, heterologous nucleic acid is nucleic acid that (if DNA encodes RNA) and proteins that are not normally produced *in vivo* by the cell in which it is expressed or that mediates or encodes mediators that alter expression of endogenous nucleic acid, such as DNA, by affecting transcription, translation, or other regulatable biochemical processes. Heterologous nucleic acid, such as DNA, can also be referred to as foreign nucleic acid, such as DNA. Any nucleic acid, such as DNA, that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which is expressed is herein encompassed by heterologous nucleic acid; heterologous nucleic acid includes exogenously added nucleic acid that is also expressed endogenously. Examples of heterologous nucleic acid include, but are not limited to, nucleic acid that encodes traceable marker proteins, such as a protein that confers drug resistance, nucleic acid that encodes therapeutically effective substances, such as anti-cancer agents, enzymes and hormones, and nucleic acid, such as DNA, that encodes other types of proteins, such as antibodies. Antibodies that are encoded by heterologous nucleic acid can be secreted or expressed on the surface of the cell in which the heterologous nucleic acid has been introduced. Heterologous nucleic acid is generally not endogenous to the cell into which it is introduced, but has been obtained from another cell or prepared synthetically. Generally, although not necessarily, such nucleic acid encodes RNA and proteins that are not normally produced by the cell in which it is expressed.

As used herein, a therapeutically effective product is a product that is encoded by heterologous nucleic acid, typically DNA, that, upon introduction of the nucleic acid into a host, a product is expressed that ameliorates or eliminates

As used herein, nucleic acid encoding a fragment or portion of an SP refers to a nucleic acid encoding only the recited fragment or portion of SP, and not the other contiguous portions of the SP.

As used herein, operative linkage of heterologous nucleic to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences refers to the relationship between such nucleic acid, such as DNA, and such sequences of nucleotides. For example, operative linkage of heterologous DNA to a promoter refers to the physical relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA in reading frame. Thus, operatively linked or operationally associated refers to the functional relationship of nucleic acid, such as DNA, with regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences. For example, operative linkage of DNA to a promoter refers to the physical and functional relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA. In order to optimize expression and/or *in vitro* transcription, it can be necessary to remove, add or alter 5' untranslated portions of the clones to eliminate extra, potential inappropriate alternative translation initiation (*i.e.*, start) codons or other sequences that can interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites (see, *e.g.*, Kozak *J. Biol. Chem. 266*:19867-19870 (1991)) can be inserted immediately 5' of the start codon and can enhance expression. The desirability of (or need for) such modification can be empirically determined.

As used herein, a sequence complementary to at least a portion of an RNA, with reference to antisense oligonucleotides, means a sequence having sufficient complementarily to be able to hybridize with the RNA, generally under moderate or high stringency conditions, forming a stable duplex; in the case of double-stranded SP antisense nucleic acids, a single strand of the duplex DNA

**TABLE 1**

| Original residue | Conservative substitution |
|---|---|
| Ala (A) | Gly; Ser, Abu |
| Arg (R) | Lys, orn |
| Asn (N) | Gln; His |
| Cys (C) | Ser |
| Gln (Q) | Asn |
| Glu (E) | Asp |
| Gly (G) | Ala; Pro |
| His (H) | Asn; Gln |
| Ile (I) | Leu; Val; Met; Nle; Nva |
| Leu (L) | Ile; Val; Met; Nle; Nv |
| Lys (K) | Arg; Gln; Glu |
| Met (M) | Leu; Tyr; Ile; NLe Val |
| Ornitine | Lys; Arg |
| Phe (F) | Met; Leu; Tyr |
| Ser (S) | Thr |
| Thr (T) | Ser |
| Trp (W) | Tyr |
| Tyr (Y) | Trp; Phe |
| Val (V) | Ile; Leu; Met; Nle; Nv |

Other substitutions are also permissible and can be determined empirically or in accord with known conservative substitutions.

As used herein, Abu is 2-aminobutyric acid; Orn is ornithine.

As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which occur in the various DNA fragments, are designated with the standard single-letter designations used routinely in the art.

As used herein, a probe or primer based on a nucleotide sequence disclosed herein, includes at least 10, 14, typically at least 16 contiguous sequence of nucleotides of SEQ ID No. 5, and probes of at least 30, 50 or 100 contiguous sequence of nucleotides of SEQ ID No. 5. The length of the probe or primer for unique hybridization is a function of the complexity of the genome of interest.

As used herein, amelioration of the symptoms of a particular disorder by administration of a particular pharmaceutical composition refers to any lessening, whether permanent or temporary, lasting or transient that can be attributed to or associated with administration of the composition.

such as by disulfide bridges. An antibody fragment generally contains at least about 50 amino acids and typically at least 200 amino acids.

As used herein, an Fv antibody fragment is composed of one variable heavy domain ($V_H$) and one variable light domain linked by noncovalent interactions.

As used herein, a dsFV refers to an Fv with an engineered intermolecular disulfide bond, which stabilizes the $V_H$-$V_L$ pair.

As used herein, an $F(ab)_2$ fragment is an antibody fragment that results from digestion of an immunoglobulin with pepsin at pH 4.0-4.5; it can be recombinantly expressed to produce the equivalent fragment.

As used herein, Fab fragments are antibody fragments that result from digestion of an immunoglobulin with papain; they can be recombinantly expressed to produce the equivalent fragment.

As used herein, scFVs refer to antibody fragments that contain a variable light chain ($V_L$) and variable heavy chain ($V_H$) covalently connected by a polypeptide linker in any order. The linker is of a length such that the two variable domains are bridged without substantial interference. Included linkers are $(Gly\text{-}Ser)_n$ residues with some Glu or Lys residues dispersed throughout to increase solubility.

As used herein, humanized antibodies refer to antibodies that are modified to include human sequences of amino acids so that administration to a human does not provoke an immune response. Methods for preparation of such antibodies are known. For example, to produce such antibodies, the hybridoma or other prokaryotic or eukaryotic cell, such as an *E. coli* or a CHO cell, that expresses the monoclonal antibody are altered by recombinant DNA techniques to express an antibody in which the amino acid composition of the non-variable region is based on human antibodies. Computer programs have been designed to identify such regions.

As used herein, diabodies are dimeric scFV; diabodies typically have shorter peptide linkers than scFvs, and they generally dimerize.

one SP, or a domain thereof, is linked, directly or indirectly via linker(s) to a targeting agent.

As used herein, a targeting agent is any moiety, such as a protein or effective portion thereof, that provides specific binding of the conjugate to a cell surface receptor, which, can internalize the conjugate or SP portion thereof. A targeting agent can also be one that promotes or facilitates, for example, affinity isolation or purification of the conjugate; attachment of the conjugate to a surface; or detection of the conjugate or complexes containing the conjugate.

As used herein, an antibody conjugate refers to a conjugate in which the targeting agent is an antibody.

As used herein, derivative or analog of a molecule refers to a portion derived from or a modified version of the molecule.

As used herein, an effective amount of a compound for treating a particular disease is an amount that is sufficient to ameliorate, or in some manner reduce the symptoms associated with the disease. Such amount can be administered as a single dosage or can be administered according to a regimen, whereby it is effective. The amount can cure the disease but, typically, is administered in order to ameliorate the symptoms of the disease. Repeated administration can be required to achieve the desired amelioration of symptoms.

As used herein equivalent, when referring to two sequences of nucleic acids means that the two sequences in question encode the same sequence of amino acids or equivalent proteins. When equivalent is used in referring to two proteins or peptides, it means that the two proteins or peptides have substantially the same amino acid sequence with only amino acid substitutions (such, as but not limited to, conservative changes such as those set forth in Table 1, above) that do not substantially alter the activity or function of the protein or peptide. When equivalent refers to a property, the property does not need to be present to the same extent (*e.g.*, two peptides can exhibit different rates of the same type of enzymatic activity), but the activities are usually substantially the same. Complementary, when referring to two nucleotide sequences, means that the two sequences of nucleotides are capable of hybridizing, typically with less than 25%, 15%, 5% or 0% mismatches between

compound is regenerated by metabolic processes. The prodrug can be designed to alter the metabolic stability or the transport characteristics of a drug, to mask side effects or toxicity, to improve the flavor of a drug or to alter other characteristics or properties of a drug. By virtue of knowledge of pharmacodynamic processes and drug metabolism *in vivo*, those of skill in this art, once a pharmaceutically active compound is known, can design prodrugs of the compound (see, *e.g.*, Nogrady (1985) *Medicinal Chemistry A Biochemical Approach*, Oxford University Press, New York, pages 388-392).

As used herein, a drug identified by the screening methods provided herein refers to any compound that is a candidate for use as a therapeutic or as a lead compound for the design of a therapeutic. Such compounds can be small molecules, including small organic molecules, peptides, peptide mimetics, antisense molecules or dsRNA, such as RNAi, antibodies, fragments of antibodies, recombinant antibodies and other such compounds that can serve as drug candidates or lead compounds.

As used herein, a peptidomimetic is a compound that mimics the conformation and certain stereochemical features of the biologically active form of a particular peptide. In general, peptidomimetics are designed to mimic certain desirable properties of a compound, but not the undesirable properties, such as flexibility, that lead to a loss of a biologically active conformation and bond breakdown. Peptidomimetics may be prepared from biologically active compounds by replacing certain groups or bonds that contribute to the undesirable properties with bioisosteres. Bioisosteres are known to those of skill in the art. For example the methylene bioisostere $CH_2S$ has been used as an amide replacement in enkephalin analogs (see, e.g., Spatola (1983) pp. 267-357 in Chemistry and Biochemistry of Amino Acids, Peptides, and Proteins, Weistein, Ed. volume 7, Marcel Dekker, New York). Morphine, which can be administered orally, is a compound that is a peptidomimetic of the peptide endorphin. For purposes herein, cyclic peptides are included among pepidomimetics.

As used herein, a promoter region or promoter element refers to a segment of DNA or RNA that controls transcription of the DNA or RNA to which it is operatively linked. The promoter region includes specific sequences that are

d) catalytic polypeptides: polymers, including polypeptides, that are capable of promoting a chemical reaction involving the conversion of one or more reactants to one or more products; such polypeptides generally include a binding site specific for at least one reactant or reaction intermediate and an active functionality proximate to the binding site, in which the functionality is capable of chemically modifying the bound reactant (see, *e.g.*, U.S. Patent No. 5,215,899);

e) hormone receptors: determination of the ligands that bind with high affinity to a receptor is useful in the development of hormone replacement therapies; for example, identification of ligands that bind to such receptors can lead to the development of drugs to control blood pressure; and

f) opiate receptors: determination of ligands that bind to the opiate receptors in the brain is useful in the development of less-addictive replacements for morphine and related drugs.

As used herein, sample refers to anything which can contain an analyte for which an analyte assay is desired. The sample can be a biological sample, such as a biological fluid or a biological tissue. Examples of biological fluids include urine, blood, plasma, serum, saliva, semen, stool, sputum, cerebral spinal fluid, tears, mucus, sperm, amniotic fluid or the like. Biological tissues are aggregate of cells, usually of a particular kind together with their intercellular substance that form one of the structural materials of a human, animal, plant, bacterial, fungal or viral structure, including connective, epithelium, muscle and nerve tissues. Examples of biological tissues also include organs, tumors, lymph nodes, arteries and individual cell(s).

As used herein: stringency of hybridization in determining percentage mismatch is as follows:

1) high stringency: 0.1 x SSPE, 0.1% SDS, 65°C
2) medium stringency: 0.2 x SSPE, 0.1% SDS, 50°C
3) low stringency: 1.0 x SSPE, 0.1% SDS, 50°C

Those of skill in this art know that the washing step selects for stable hybrids and also know the ingredients of SSPE (see, *e.g.*, Sambrook, E.F. Fritsch, T. Maniatis, in: *Molecular Cloning, A Laboratory Manual*, Cold Spring

$^{32}$P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 hours at 55°C, and then washed twice for 30 minutes at 60°C in a solution containing 1X SSC and 0.1% SDS. Filters are blotted dry and exposed for autoradiography. Other conditions of moderate stringency which can be used are well-known in the art. Washing of filters is done at 37°C for 1 hour in a solution containing 2X SSC, 0.1% SDS.

By way of example and not way of limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 hours to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 μg/ml denatured salmon sperm DNA. Filters are hybridized for 48 hours at 65°C in prehybridization mixture containing 100 μg/ml denatured salmon sperm DNA and 5-20 X $10^6$ cpm of $^{32}$P-labeled probe. Washing of filters is done at 37°C for 1 hour in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA. This is followed by a wash in 0.1X SSC at 50°C for 45 minutes before autoradiography. Other conditions of high stringency which can be used are well known in the art.

The term substantially identical or substantially homologous or similar varies with the context as understood by those skilled in the relevant art and generally means at least 60% or 70%, preferably means at least 80%, 85% or more preferably at least 90%, and most preferably at least 95% identity.

As used herein, substantially identical to a product means sufficiently similar so that the property of interest is sufficiently unchanged so that the substantially identical product can be used in place of the product.

As used herein, substantially pure means sufficiently homogeneous to appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and biological activities, of the substance. Methods for purification of the compounds to produce substantially chemically pure compounds are known to

host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those of skill in the art and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome.

As used herein, protein binding sequence refers to a protein or peptide sequence that is capable of specific binding to other protein or peptide sequences generally, to a set of protein or peptide sequences or to a particular protein or peptide sequence.

As used herein, epitope tag refers to a short stretch of amino acid residues corresponding to an epitope to facilitate subsequent biochemical and immunological analysis of the epitope tagged protein or peptide. Epitope tagging is achieved by including the sequence of the epitope tag to the protein-encoding sequence in an appropriate expression vector. Epitope tagged proteins can be affinity purified using highly specific antibodies raised against the tags.

As used herein, metal binding sequence refers to a protein or peptide sequence that is capable of specific binding to metal ions generally, to a set of metal ions or to a particular metal ion.

As used herein, a combination refers to any association between two or among more items.

As used herein, a composition refers to a any mixture. It can be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

As used herein, fluid refers to any composition that can flow. Fluids thus encompass compositions that are in the form of semi-solids, pastes, solutions, aqueous mixtures, gels, lotions, creams and other such compositions.

As used herein, a cellular extract refers to a preparation or fraction which is made from a lysed or disrupted cell.

As used herein, an agent is said to be randomly selected when the agent is chosen randomly without considering the specific sequences involved in the association of a protein alone or with its associated substrates, binding partners, etc. An example of randomly selected agents is the use a chemical library or a

For purposes herein, the protease domain of the MTSP does not have to result from activation cleavage, which produces a two chain activated product, but rather includes single chain polypeptides where the N-terminii include the consensus sequence ↓VVGG, ↓IVGG, ↓VGLL, ↓ILGG, ↓IVQG or ↓IVNG ↓IASG or other such motif. Such polypeptides, although not the result of activation cleavage and not two-chain forms, exhibit proteolytic (catalytic) activity. These protease domain polypeptides are used in assays to screen for agents that modulate the activity of the MTSP9.

The MTSP family is a target for therapeutic intervention and also some members can serve as diagnostic markers for tumor development, growth and/or progression. As discussed, the members of this family are involved in proteolytic processes that are implicated in tumor development, growth and/or progression. This implication is based upon their functions as proteolytic enzymes in processes related to ECM degradation and/or remodeling and activation of pro-growth factors, pro-hormones or pro-angiogenic compounds. In addition, their levels of expression or level of activation or their apparent activity resulting from substrate levels or alterations in substrates and levels thereof differs in tumor cells and non-tumor cells in the same tissue. Similarly the level of co-factors or receptors for these proteases can vary between tumor and non-tumor cells. Hence, protocols and treatments that alter their activity, such as their proteolytic activities and roles in signal transduction, and/or their expression, such as by contacting them with a compound that modulates their activity and/or expression, could impact tumor development, growth and/or progression. Also, in some instances, the level of activation and/or expression can be altered in tumors, such as lung carcinoma, colon adenocarcinoma and ovarian carcinoma.

**MTSP9**

MTSP9 is of interest because it is expressed or is active in tumor cells. The MTSP provided herein can serve as a diagnostic marker for particular tumors, by virtue of a level of activity and/or expression or function in a subject (i.e. a mammal, particularly a human) with neoplastic disease, compared to a subject or subjects that do not have the neoplastic disease. In addition,

Exemplary MTSP9-encoding nucleic acid and protein sequences of a protease domain are set forth in SEQ ID Nos. 5 and 6, and a full-length protein and encoding nucleic acid sequence is set SEQ ID Nos. 18 and 17. Thus, an MTSP9 polypeptide includes the sequence of amino acids set forth in SEQ ID Nos. 6, 16 or 18. Smaller portions thereof that retain protease activity are contemplated. The protease domain thereof is set forth in SEQ ID No. 16.

Substantially purified MTSP9 protease is encoded by a nucleic acid that hybridizes to a nucleic acid molecule containing the protease domain encoded by the nucleotide sequence set forth in any of SEQ. ID Nos. 5 and 17 under at least moderate, generally high, stringency conditions, such that the protease domain encoding nucleic acid thereof hybridizes along its full-length or at least 70%, 80% or 90% of the full-length. In certain embodiments the substantially purified MTSP protease is a single chain polypeptide that includes substantially the sequence of amino acids set forth in SEQ ID No. 6, 18 or the protease domain portion thereof, or a catalytically active portion thereof.

Also included are substantially purified MTSP9 zymogens, activated two chain forms, single chain protease domains and two chain protease domains. These polypeptides are encoded by a nucleic acid that includes sequence encoding a protease domain that exhibits proteolytic activity and that hybridizes to a nucleic acid molecule having a nucleotide sequence set forth in SEQ ID No. 5 or 7, typically under moderate, generally under high stringency, conditions and generally along the full-length or along at least about 70%, 80% or 90% of the full-length (or substantially the full-length) of the protease domain. Splice variants are also contemplated herein.

**Protease domains**

MTSP protease domains include the single chain protease domains of MTSP9. Provided are the protease domains or proteins that include a portion of an MTSP that is the protease domain of any MTSP, particularly a MTSP9. The protein can also include other non-MTSP sequences of amino acids, but includes the protease domain or a sufficient portion thereof to exhibit catalytic activity in any *in vitro* assay that assess such protease activity, such as any provided

the sequence resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent alteration. Substitutes for an amino acid within the sequence can be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid (see, *e.g.*, Table 1). Muteins of the MTSP9 or a domain thereof, such as a protease domain, in which up to about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90% or 95% of the amino acids are replaced with another amino acid are provided. Generally such muteins retain at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% of the protease activity the unmutated protein.

Muteins of the MTSP9 or a domain thereof, such as a protease domain, in which up to about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% of the amino acids are replaced with another amino acid are provided. Generally such muteins retain at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% of the protease activity the unmutated protein.

Included among the polypeptides provided herein are the MTSP9 protease domain or a polypeptide with amino acid changes such that the specificity and protease activity remains substantially unchanged or changed (increased or decreased) by a specified percentage, such as 10, 20, 30, 40, 50%. In particular, a substantially purified mammalian MTSP polypeptide is provided that has a transmembrane domain and can additionally include a transmembrane (TM) domain, a SEA domain and a serine protease catalytic domain is provided.

Also provided is a substantially purified protein containing a sequence of amino acids that has at least 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%,

Muteins can be made by making conservative amino acid substitions and also non-conservative amino acid substitutions. For example, amino acid substitutions the desirably alter properties of the proteins can be made. In one embodiment, mutations that prevent degradation of the polypeptide can be made. Many proteases cleave after basic residues, such as R and K; to eliminate such cleavage, the basic residue is replaced with a non-basic residue. Also, non-conservative changes at amino acids outside of the protease domain can be effected without altering protease activity. Non-conservative changes at amino acids that are responsible for activities other than protease activity may be desirable. For example, interaction of the protease with an inhibitor can be blocked while retaining catalytic activity by effecting a non-conservative change at the site interaction of the inhibitor with the protease. Similarly, receptor binding can be altered without altering catalytic activity by effecting a non-conservative or conservative at a site of interaction of the receptor with the protease.

Antigenic epitopes that contain at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 40, 50, and typically 10-15 amino acids of the MTSp9 polypeptide are provided. These antigenic epitopes are used, for example, to raise antibodies. Antibodies specific for each epitope or combinations thereof and for single and two-chain forms are also provided.

**Nucleic acid molecules, vectors and plasmids, cells and expression of MTSP9 polyeptides**

**Nucleic acid molecules**

Due to the degeneracy of nucleotide coding sequences, other nucleic sequences which encode substantially the same amino acid sequence as a MTSP are contemplated. These include but are not limited to nucleic acid molecules that include all or portions of MTSP9-encoding genes that are altered by the substitution of different codons that encode the amino acid residue within the sequence, thus producing a silent change.

**Nucleic acids**

Also provided herein are nucleic acid molecules that encode MTSP9 polypeptides and the encoded proteins. In particular, nucleic acid molecules

the muteins are those in which the Cys residue-encoding codons, is/are replaced with other amino acid residues, such as a codon encoding a serine. Such muteins are also provided herein. Each of such domains is provided herein as are nucleic acid molecules that include sequences of nucleotides encoding such domains. Some MTSPs can additionally include a LDLR domain, a scavenger-receptor cysteine rich (SRCR) domain and other domains.

The isolated nucleic acid fragment is DNA, including genomic or cDNA, or is RNA, or can include other components, such as protein nucleic acid and other nucleotide analogs. The isolated nucleic acid can include additional components, such as heterologous or native promoters, and other transcriptional and translational regulatory sequences, these genes can be linked to other genes, such as reporter genes or other indicator genes or genes that encode indicators.

Also provided are nucleic acid molecules that hybridize to the above-noted sequences of nucleotides encoding MTSP9 at least at low stringency, moderate stringency, and typically at high stringency, and that encode the protease domain and/or the full-length protein or at least 70%, 80% or 90% of the full-length protease domain or other domains of an MTSP9 or a splice variant or allelic variant thereof. Generally the molecules hybridize under such conditions along their full-length or along at least 70%, 80% or 90% of the full-length for at least one domain and encode at least one domain, such as the protease or extracellular domain, of the polypeptide. In particular, such nucleic acid molecules include any isolated nucleic fragment that encodes at least one domain of a membrane serine protease, that (1) contains a sequence of nucleotides that encodes the protease or a domain thereof, and (2) is selected from among:

(a) a sequence of nucleotides that encodes the protease or a domain thereof that includes a sequence of nucleotides set forth in SEQ ID Nos. 15 or 17;

(b) a sequence of nucleotides that encodes such portion or the full-length protease and hybridizes under conditions of high stringency, generally to nucleic acid that is complementary to a mRNA

at least about 16 nucleotides, often at least about 30 nucleotides. The length of the probe or primer is a function of the size of the genome probed; the larger the genome, the longer the probe or primer required for specific hybridization to a single site. Those of skill in the art can select appropriately sized probes and primers. Generally probes and primers as described are single-stranded. Double stranded probes and primers can be used, if they are denatured when used.

Probes and primers derived from the nucleic acid molecules are provided. Such probes and primers contain at least 8, 14, 16, 30, 100 or more contiguous nucleotides with identity to contiguous nucleotides of an MTSP9, generally, except for nucleotides 634-751 of SEQ ID No. 5 or nucleotides 1162-1279 of SEQ ID No. 17, and probes of at least 30, 50 or 100 contiguous sequence of nucleotides of SEQ ID No. 5, except for nucleotides 634-734 of SEQ ID No. 5 (except for nucleotides 1162-1262 of SEQ ID No. 18). The probes and primers are optionally labelled with a detectable label, such as a radiolabel or a fluorescent tag, or can be mass differentiated for detection by mass spectrometry or other means.

Also provided is an isolated nucleic acid molecule that includes the sequence of molecules that is complementary to the nucleotide sequence encoding MTSP9 or the portion thereof. Double-stranded RNA (dsRNA), such as RNAi is also provided.

**Plasmids, Vectors and Cells**

Plasmids and vectors containing the nucleic acid molecules are also provided. Cells containing the vectors, including cells that express the encoded proteins are provided. The cell can be a bacterial cell, a yeast cell, a fungal cell, a plant cell, an insect cell or an animal cell. Methods for producing an MTSP or single chain form of the protease domain thereof by, for example, growing the cell under conditions whereby the encoded MTSP is expressed by the cell, and recovering the expressed protein, are provided herein. As noted, for MTSP9, the full-length zymogens and activated proteins and activated (two chain) protease

blood leukocyte, lung (adult and fetal), pancreas, lymph node, bone marrow, trachea, uterus, prostate, testes, ovary and the gland organs (mammary, adrenal, thyroid, pituitary and salivary). MTSP9 is also expressed in esophagus tumor tissues, in a lung carcinoma and, at a lower level, in a colorectal carcinoma, lymphoma, a cervical carcinoma (HeLaS3) and leukemia cell lines.

**D. Identification and isolation of MTSP9 polypeptide genes**

The MTSP polypeptides and/or domains thereof, can be obtained by methods well known in the art for protein purification and recombinant protein expression. Any method known to those of skill in the art for identification of nucleic acids that encode desired genes can be used. Any method available in the art can be used to obtain a full-length (*i.e.*, encompassing the entire coding region) cDNA or genomic DNA clone encoding an MTSP polypeptide. For example, the polymerase chain reaction (PCR) can be used to amplify a sequence that is expressed in normal and tumor cells or tissues, *e.g.*, nucleic acids encoding an MTSP9 polypeptide (SEQ. Nos: 5 and 17), in a genomic or cDNA library. Oligonucleotide primers that hybridize to sequences at the 3' and 5' termini of the identified sequences can be used as primers to amplify by PCR sequences from a nucleic acid sample (RNA or DNA), generally a cDNA library, from an appropriate source (*e.g.*, tumor or cancer tissue).

PCR can be carried out, *e.g.*, by use of a Perkin-Elmer Cetus thermal cycler and Taq polymerase (Gene Amp™). The DNA being amplified can include mRNA or cDNA or genomic DNA from any eukaryotic species. One can choose to synthesize several different degenerate primers, for use in the PCR reactions. It is also possible to vary the stringency of hybridization conditions used in priming the PCR reactions, to amplify nucleic acid homologs (*e.g.*, to obtain *MTSP polypeptide* sequences from species other than humans or to obtain human sequences with homology to MTSP9 polypeptide) by allowing for greater or lesser degrees of nucleotide sequence similarity between the known nucleotide sequence and the nucleic acid homolog being isolated. For cross-species hybridization, low stringency to moderate stringency conditions are used. For same species hybridization, moderately stringent to highly stringent conditions are used. The conditions can be empirically determined.

The DNA can be cleaved at specific sites using various restriction enzymes. Alternatively, one can use DNAse in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, for example, by sonication. The linear DNA fragments then can be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography.

Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired gene can be accomplished in a number of ways. For example, a portion of the MTSP polypeptide (of any species) gene (*e.g.*, a PCR amplification product obtained as described above or an oligonucleotide having a sequence of a portion of the known nucleotide sequence) or its specific RNA, or a fragment thereof be purified and labeled, and the generated DNA fragments can be screened by nucleic acid hybridization to the labeled probe (Benton and Davis, *Science* 196:180 (1977); Grunstein and Hogness, *Proc. Natl. Acad. Sci. U.S.A.* 72:3961 (1975)). Those DNA fragments with substantial homology to the probe will hybridize. It is also possible to identify the appropriate fragment by restriction enzyme digestion(s) and comparison of fragment sizes with those expected according to a known restriction map if such is available or by DNA sequence analysis and comparison to the known nucleotide sequence of MTSP polypeptide. Further selection can be carried out on the basis of the properties of the gene. Alternatively, the presence of the gene can be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones, or DNA clones which hybrid-select the proper mRNA, can be selected which produce a protein that, *e.g.*, has similar or identical electrophoretic migration, isolectric focusing behavior, proteolytic digestion maps, antigenic properties, serine protease activity. If an anti-MTSP polypeptide antibody is available, the protein can be identified by binding of labeled antibody to the putatively MTSP polypeptide synthesizing clones, in an ELISA (enzyme-linked immunosorbent assay)-type procedure.

Alternatives to isolating the MTSP9 polypeptide genomic DNA include, but are not limited to, chemically synthesizing the gene sequence from a known

MTSP polypeptide can be inserted into an appropriate expression vector, *i.e.*, a vector that contains the necessary elements for the transcription and translation of the inserted protein coding sequence. The necessary transcriptional and translational signals can also be supplied by the native promoter for MTSP genes, and/or their flanking regions.

Also provided are vectors that contain nucleic acid encoding the MTSPs. Cells containing the vectors are also provided. The cells include eukaryotic and prokaryotic cells, and the vectors are any suitable for use therein.

Prokaryotic and eukaryotic cells, including endothelial cells, containing the vectors are provided. Such cells include bacterial cells, yeast cells, fungal cells, plant cells, insect cells and animal cells. The cells are used to produce an MTSP polypeptide or protease domain thereof by (a) groowing the above-described cells under conditions whereby the encoded MTSP polypeptide or protease domain of the MTSP polypeptide is expressed by the cell, and then (b) recovering the expressed protease domain protein. In the exemplified embodiments, the protease domain is secreted into the medium.

In one embodiment, the vectors include a sequence of nucleotides that encodes a polypeptide that has protease activity and contains all or a portion of only the protease domain, or multiple copies thereof, of an SP protein are provided. Also provided are vectors that comprise a sequence of nucleotides that encodes the protease domain and additional portions of an SP protein up to and including a full length SP protein, as well as multiple copies thereof, are also provided. The vectors can selected for expression of the SP protein or protease domain thereof in the cell or such that the SP protein is expressed as a secreted protein. Alternatively, the vectors can include signals necessary for secretion of encoded proteins. When the protease domain is expressed the nucleic acid is linked to nucleic acid encoding a secretion signal, such as the *Saccharomyces cerevisiae* $\alpha$ mating factor signal sequence or a portion thereof, or the native signal sequence.

A variety of host-vector systems can be used to express the protein coding sequence. These include but are not limited to mammalian cell systems infected with virus (*e.g.* vaccinia virus, adenovirus, etc.); insect cell systems

alcohol dehydrogenase promoter, the phosphoglycerol kinase promoter, the alkaline phosphatase promoter, and the following animal transcriptional control regions that exhibit tissue specificity and have been used in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift et al., *Cell* 38:639-646 (1984); Ornitz et al., *Cold Spring Harbor Symp. Quant. Biol.* 50:399-409 (1986); MacDonald, *Hepatology* 7:425-515 (1987)); insulin gene control region which is active in pancreatic beta cells (Hanahan et al., *Nature* 315:115-122 (1985)), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., *Cell* 38:647-658 (1984); Adams et al., *Nature* 318:533-538 (1985); Alexander et al., *Mol. Cell Biol.* 7:1436-1444 (1987)), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder et al., *Cell* 45:485-495 (1986)), albumin gene control region which is active in liver (Pinckert et al., *Genes and Devel.* 1:268-276 (1987)), alpha-fetoprotein gene control region which is active in liver (Krumlauf et al., *Mol. Cell. Biol.* 5:1639-1648 (1985); Hammer et al., *Science* 235:53-58 1987)), alpha-1 antitrypsin gene control region which is active in liver (Kelsey et al., *Genes and Devel.* 1:161-171 (1987)), beta globin gene control region which is active in myeloid cells (Mogram et al., *Nature* 315:338-340 (1985); Kollias et al., *Cell* 46:89-94 (1986)), myelin basic protein gene control region which is active in oligodendrocyte cells of the brain (Readhead et al., *Cell* 48:703-712 (1987)), myosin light chain-2 gene control region which is active in skeletal muscle (Sani, *Nature* 314:283-286 (1985)), and gonadotrophic releasing hormone gene control region which is active in gonadotrophs of the hypothalamus (Mason et al., *Science* 234:1372-1378 (1986)).

In a specific embodiment, a vector is used that contains a promoter operably linked to nucleic acids encoding an MTSP polypeptide, or a domain, fragment, derivative or homolog, thereof, one or more origins of replication, and optionally, one or more selectable markers (*e.g.*, an antibiotic resistance gene). Expression vectors containing the coding sequences, or portions thereof, of an MTSP polypeptide, is made, for example, by subcloning the coding portions into the EcoRI restriction site of each of the three pGEX vectors (glutathione S-transferase expression vectors (Smith and Johnson, *Gene* 7:31-40 (1988)). This

The MTSP polypeptides can be isolated and purified by standard methods known in the art (either from natural sources or recombinant host cells expressing the complexes or proteins), including but not restricted to column chromatography (*e.g.*, ion exchange, affinity, gel exclusion, reversed-phase high pressure and fast protein liquid), differential centrifugation, differential solubility, or by any other standard technique used for the purification of proteins. Functional properties can be evaluated using any suitable assay known in the art.

Alternatively, once an MTSP polypeptide or its domain or derivative is identified, the amino acid sequence of the protein can be deduced from the nucleotide sequence of the gene which encodes it. As a result, the protein or its domain or derivative can be synthesized by standard chemical methods known in the art (*e.g.* see Hunkapiller et al, *Nature* 310:105-111 (1984)).

Manipulations of MTSP polypeptide sequences can be made at the protein level. Also contemplated herein are MTSP polypeptide proteins, domains thereof, derivatives or analogs or fragments thereof, which are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand. Any of numerous chemical modifications can be carried out by known techniques, including but not limited to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, $NaBH_4$, acetylation, formylation, oxidation, reduction, metabolic synthesis in the presence of tunicamycin and other such agents.

In addition, domains, analogs and derivatives of an MTSP polypeptide can be chemically synthesized. For example, a peptide corresponding to a portion of an MTSP polypeptide, which includes the desired domain or which mediates the desired activity *in vitro* can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the MTSP polypeptide sequence. Non-classical amino acids include but are not limited to the D-isomers of the common amino acids, a-amino isobutyric acid, 4-aminobutyric acid, Abu,

glycosylation sites are contemplated; the MTSP9 of SEQ ID No. 18 has two potential glycosylation sites. Such mutations can be effected by any technique for mutagenesis known in the art, including, but not limited to, chemical mutagenesis and *in vitro* site-directed mutagenesis (Hutchinson et al., *J. Biol. Chem.* 253:6551-6558 (1978)), use of TAB® linkers (Pharmacia). In one embodiment, for example, an MTSP polypeptide or domain thereof is modified to include a fluorescent label. In other specific embodiments, the MTSP polypeptide is modified to have a heterobifunctional reagent, such heterobifunctional reagents can be used to crosslink the members of the complex.

In addition, domains, analogs and derivatives of an MTSP can be chemically synthesized. For example, a peptide corresponding to a portion of an MTSP, which includes the desired domain or which mediates the desired activity *in vitro* can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the MTSP sequence. Non-classical amino acids include but are not limited to the D-isomers of the common amino acids, a-amino isobutyric acid, 4-aminobutyric acid, Abu, 2-aminobutyric acid, $\epsilon$-Abu, e-Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionoic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, ß-alanine, fluoro-amino acids, designer amino acids such as ß-methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

**F. SCREENING METHODS**

The single chain protease domains, as shown herein, can be used in a variety of methods to identify compounds that modulate the activity thereof. For SPs that exhibit higher activity or expression in tumor cells, compounds that inhibit the proteolytic activity are of particular interest. For any SPs that are active at lower levels in tumor cells, compounds or agents that enhance the

In one embodiment a plurality of the test substances are screened simultaneously in the above screening method. In another embodiment, the MTSP9 is isolated from a target cell as a means for then identifying agents that are potentially specific for the target cell.

In another embodiment, a test substance is a therapeutic compound, and whereby a difference of the MTSP9 activity measured in the presence and in the absence of the test substance indicates that the target cell responds to the therapeutic compound.

One method includes the steps of (a) contacting the MTSP9 polypeptide or protease domain thereof with one or a plurality of test compounds under conditions conducive to interaction between the ligand and the compounds; and (b) identifying one or more compounds in the plurality that specifically binds to the ligand.

Another method provided herein includes the steps of a) contacting an MTSP9 polypeptide or protease domain thereof with a substrate of the MTSP9 polypeptide, and detecting the proteolysis of the substrate, whereby the activity of the MTSP9 polypeptide is assessed; b) contacting the MTSP9 polypeptide with a substrate of the MTSP9 polypeptide in the presence of a test substance, and detecting the proteolysis of the substrate, whereby the activity of the MTSP9 polypeptide is assessed; and c) comparing the activity of the MTSP9 polypeptide assessed in steps a) and b), whereby the activity measured in step a) differs from the activity measured in step b) indicates that the test substance modulates the activity of the MTSP9 polypeptide.

In another embodiment, a plurality of the test substances are screened simultaneously. In comparing the activity of an MTSP9 polypeptide in the presence and absence of a test substance to assess whether the test substance is a modulator of the MTSP9 polypeptide, it is unnecessary to assay the activity in parallel, although such parallel measurement is typical. It is possible to measure the activity of the MTSP9 polypeptide at one time point and compare the measured activity to a historical value of the activity of the MTSP9 polypeptide.

methods include any MTSP9 polypeptide as defined herein, including the MTSP9 single chain protease domain or proteolytically active portion thereof.

A variety of methods are provided herein. These methods can be performed in solution or in solid phase reactions in which the MTSP9 polypeptide(s) or protease domain(s) thereof are linked, either directly or indirectly via a linker, to a solid support. Screening assays are described in the Examples, and these assays have been used to identify candidate compounds. For purposes herein, all binding assays described above are provided for MTSP9.

Methods for identifying an agent, such as a compound, that specifically binds to an MTSP9 single chain protease domain, a zymogen or full-length activated MTSP9 or two chain protease domain thereof are provided herein. The method can be practiced by (a) contacting the MTSP9 with one or a plurality of test agents under conditions conducive to binding between the MTSP9 and an agent; and (b) identifying one or more agents within the plurality that specifically binds to the MTSP9.

For example, in practicing such methods the MTSP9 polypeptide is mixed with a potential binding partner or an extract or fraction of a cell under conditions that allow the association of potential binding partners with the polypeptide. After mixing, peptides, polypeptides, proteins or other molecules that have become associated with an MTSP9 are separated from the mixture. The binding partner that bound to the MTSP9 can then be removed and further analyzed. To identify and isolate a binding partner, the entire protein, for instance the entire disclosed protein of SEQ ID Nos. 6 can be used. Alternatively, a fragment of the protein can be used.

A variety of methods can be used to obtain cell extracts or body fluids, such as blood, serum, urine, sweat, synovial fluid, CSF and other such fluids. For example, cells can be disrupted using either physical or chemical disruption methods. Examples of physical disruption methods include, but are not limited to, sonication and mechanical shearing. Examples of chemical lysis methods include, but are not limited to, detergent lysis and enzyme lysis. A skilled artisan

Another *in vitro* binding assay, particularly for an MTSP9, uses a mixture of a polypeptide that contains at least the catalytic domain of one of these proteins and one or more candidate binding targets or substrates. After incubating the mixture under appropriate conditions, the ability of the MTSP9 or a polypeptide fragment thereof containing the catalytic domain to bind to or interact with the candidate substrate is assessed. For cell-free binding assays, one of the components includes or is coupled to a detectable label. The label can provide for direct detection, such as radioactivity, luminescence, optical or electron density, *etc.*, or indirect detection such as an epitope tag, an enzyme, *etc*. A variety of methods can be employed to detect the label depending on the nature of the label and other assay components. For example, the label can be detected bound to the solid substrate or a portion of the bound complex containing the label can be separated from the solid substrate, and the label thereafter detected.

**3. Detection of signal transduction**

MTSP9, which is a transmembrane protein, can be involved directly or indirectly in signal transduction directly as a cell surface receptor or indirectly by activating proteins, such as pro-growth factors that can initiate signal transduction.

In addition, secreted of MTSP9, such as the extracellular domain of MTSP9, can be involved in signal transduction either directly by binding to or interacting with a cell surface receptor or indirectly by activating proteins, such as pro-growth factors that can initiate signal transduction. Assays for assessing signal transduction are well known to those of skill in the art, and can be adapted for use with the MTSP9 polypeptide.

Assays for identifying agents that affect or alter signal transduction mediated directly or indirectly, such as via activation of a pro-growth factor, by an MTSP9, particularly the full length or a sufficient portion to anchor the extracellular domain or a functional portion thereof of an MTSP9 on the surface of a cell are provided. Such assays, include, for example, transcription based assays in which modulation of a transduced signal is assessed by detecting an

be chosen to maximize the difference in stability between the probe:target hybrid and potential probe:non-target hybrids.

Probes can be designed from the nucleic acids through methods known in the art. For instance, the G+C content of the probe and the probe length can affect probe binding to its target sequence. Methods to optimize probe specificity are commonly available (see, *e.g.*, Sambrook *et al.* (1989) MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed. Cold Spring Harbor Laboratory Press); and Ausubel *et al.* (1995) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Greene Publishing Co., NY).

Hybridization conditions are modified using known methods (see, *e.g.*, Sambrook *et al.* (1989) MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed. Cold Spring Harbor Laboratory Press); and Ausubel *et al.* (1995) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Greene Publishing Co., NY), as required for each probe. Hybridization of total cellular RNA or RNA enriched for polyA RNA can be accomplished in any available format. For instance, total cellular RNA or RNA enriched for polyA RNA can be affixed to a solid support, and the solid support exposed to at least one probe comprising at least one, or part of one of the nucleic acid molecules under conditions in which the probe specifically hybridizes. Alternatively, nucleic acid fragments comprising at least one, or part of one of the sequences can be affixed to a solid support, such as a porous glass wafer. The glass wafer can then be exposed to total cellular RNA or polyA RNA from a sample under conditions in which the affixed sequences specifically hybridize. Such glass wafers and hybridization methods are widely available, for example, those disclosed by Beattie (WO 95/11755). By examining for the ability of a given probe to specifically hybridize to an RNA sample from an untreated cell population and from a cell population exposed to the agent, agents which up or down regulate the expression of a nucleic acid encoding the MTSP9 polypeptide, are identified.

extended at either the amino or carboxy terminus with a Cys residue or interspersed with cysteine residues, for example, to facilitate linking to a carrier. Administration of the immunogens is conducted generally by injection over a suitable time period and with use of suitable adjuvants, as is generally understood in the art. During the immunization schedule, titers of antibodies are taken to determine adequacy of antibody formation.

Anti-peptide antibodies can be generated using synthetic peptides corresponding to, for example, the carboxy terminal amino acids of the MTSP9. Synthetic peptides can be as small as 1-3 amino acids in length, generally at least 4 or more amino acid residues long. The peptides can be coupled to KLH using standard methods and can be immunized into animals, such as rabbits or ungulates. Polyclonal antibodies can then be purified, for example using Actigel beads containing the covalently bound peptide.

While the polyclonal antisera produced in this way can be satisfactory for some applications, for pharmaceutical compositions, use of monoclonal preparations are generally used. Immortalized cell lines which secrete the desired monoclonal antibodies can be prepared using the standard method of Kohler *et al.*, (*Nature* 256: 495-7 (1975)) or modifications which effect immortalization of lymphocytes or spleen cells, as is generally known. The immortalized cell lines secreting the desired antibodies are screened by immunoassay in which the antigen is the peptide hapten, polypeptide or protein. When the appropriate immortalized cell culture secreting the desired antibody is identified, the cells can be cultured either *in vitro* or by production *in vivo* via ascites fluid. Of particular interest, are monoclonal antibodies that recognize the catalytic domain or activation cleavage site (region) of an MTSP9.

Additionally, the zymogen or two-chain form of the MTSP9 can be used to make monoclonal antibodies that recognize conformation epitopes. The desired monoclonal antibodies are then recovered from the culture supernatant or from the ascites supernatant. Fragments of the monoclonals or the polyclonal antisera which contain the immunologically significant portion can be used as antagonists, as well as the intact antibodies. Use of immunologically reactive fragments, such as the Fab, Fab', of $F(ab')_2$ fragments are often used, especially

of the test substances are screened simultaneously (*See generally, High Throughput Screening: The Discovery of Bioactive Substances* (Devlin, Ed.) Marcel Dekker, 1997; Sittampalam et al., *Curr. Opin. Chem. Biol.*, *1*:384-91 (1997); and Silverman et al., *Curr. Opin. Chem. Biol.*, *2*:397-403 (1998)). For example, the assay can be conducted in a multi-well (*e.g.*, 24-, 48-, 96-, 384-, 1536-well or higher density), chip or array format.

High-throughput screening (HTS) is the process of testing a large number of diverse chemical structures against disease targets to identify "hits" (Sittampalam et al., *Curr. Opin. Chem. Biol.*, *1*:384-91 (1997)). Current state-of-the-art HTS operations are highly automated and computerized to handle sample preparation, assay procedures and the subsequent processing of large volumes of data.

Detection technologies employed in high-throughput screens depend on the type of biochemical pathway being investigated (Sittampalam et al., *Curr. Opin. Chem. Biol.*, *1*:384-91 (1997)). These methods include, radiochemical methods, such as the scintillation proximity assays (SPA), which can be adapted to a variety of enzyme assays (Lerner et al., *J. Biomol. Screening*, *1*:135-143 (1996); Baker et al., *Anal. Biochem.*, *239*:20-24 (1996); Baum et al., *Anal. Biochem.*, *237*:129-134 (1996); and Sullivan et al., *J. Biomol. Screening* *2*:19-23 (1997)) and protein-protein interaction assays (Braunwalder et al., *J. Biomol. Screening* *1*:23-26 (1996); Sonatore et al., *Anal. Biochem. 240*:289-297 (1996); and Chen et al., *J. Biol. Chem. 271*:25308-25315 (1996)), and non-isotopic detection methods, including but are not limited to, colorimetric and luminescence detection methods, resonance energy transfer (RET) methods, time-resolved fluorescence (HTRF) methods, cell-based fluorescence assays, such as fluorescence resonance energy transfer (FRET) procedures (see, *e.g.*,Gonzalez et al., *Biophys. J.*, *69*:1272-1280 (1995)), fluorescence polarization or anisotropy methods (see, *e.g.*, Jameson et al., *Methods Enzymol. 246*:283-300 (1995); Jolley, *J. Biomol. Screening* *1*:33-38 (1996); Lynch et al., *Anal. Biochem. 247*:77-82 (1997)), fluorescence correlation spectroscopy (FCS) and other such methods.

compound that demonstrates a low $IC_{50}$ value or $EC_{50}$ value for the target enzyme, *e.g.*, MTSP9 polypeptide, and a higher $IC_{50}$ value or $EC_{50}$ value for other enzymes within the test panel (*e.g.*, urokinase tissue plasminogen activator, thrombin, Factor Xa), is considered to be selective toward the target enzyme. Generally, a compound is deemed selective if its $IC_{50}$ value or $EC_{50}$ value in the target enzyme assay is at least one order of magnitude less than the next smallest $IC_{50}$ value or $EC_{50}$ value measured in the selectivity panel of enzymes.

Compounds are also evaluated for their activity *in vivo*. The type of assay chosen for evaluation of test compounds depends on the pathological condition to be treated or prevented by use of the compound, as well as the route of administration to be evaluated for the test compound.

For instance, to evaluate the activity of a compound to reduce tumor growth through inhibition of MTSP9 polypeptide, the procedures described by Jankun et al., *Canc. Res. 57*:559-563 (1997) to evaluate PAI-1 can be employed. Briefly, the ATCC cell lines DU145 and LnCaP are injected into SCID mice. After tumors are established, the mice are given test compound according to a dosing regime determined from the compound's *in vitro* characteristics. The Jankun *et al.* compound was administered in water. Tumor volume measurements are taken twice a week for about five weeks. A compound is deemed active if an animal to which the compound was administered exhibited decreased tumor volume, as compared to animals receiving appropriate control compounds.

Another *in vivo* experimental model designed to evaluate the effect of p-aminobenzamidine, a swine protease inhibitor, on reducing tumor volume is described by Billström et al., *Int. J. Cancer 61*:542-547 (1995).

To evaluate the ability of a compound to reduce the occurrence of, or inhibit, metastasis, the procedures described by Kobayashi *et al. Int. J. Canc. 57*:727-733d (1994) can be employed. Briefly, a murine xenograft selected for high lung colonization potential in injected into C57B1/6 mice i.v. (experimental metastasis) or s.c. into the abdominal wall (spontaneous metastasis). Various concentrations of the compound to be tested can be admixed with the tumor cells in Matrigel prior to injection. Daily i.p. injections of the test compound are

which neovascularization can be visualized, are photographed. An experimental animal receiving Matrigel and an effective dose of test compound exhibits less vascularization than a control animal or an experimental animal receiving a less- or non-effective does of compound.

An *in vivo* system designed to test compounds for their ability to limit the spread of primary tumors is described by Crowley et al., *Proc. Natl. Acad. Sci. 90*:5021-5025 (1993). Nude mice are injected with tumor cells (PC3) engineered to express CAT (chloramphenicol acetyltransferase). Compounds to be tested for their ability to decrease tumor size and/or metastases are administered to the animals, and subsequent measurements of tumor size and/or metastatic growths are made. In addition, the level of CAT detected in various organs provides an indication of the ability of the test compound to inhibit metastasis; detection of less CAT in tissues of a treated animal versus a control animal indicates less CAT-expressing cells migrated to that tissue.

*In vivo* experimental modes designed to evaluate the inhibitory potential of a test serine protease inhibitors, using a tumor cell line F3II known to be highly invasive (see, *e.g.*, Alonso et al., *Breast Canc. Res. Treat. 40*:209-223 (1996)). Alonso describes *in vivo* studies for toxicity determination, tumor growth, invasiveness, spontaneous metastasis, experimental lung metastasis, and an angiogenesis assay.

The CAM model (chick embryo chorioallantoic membrane model), first described by L. Ossowski in 1998 (*J. Cell Biol. 107*:2437-2445 (1988)), provides another method for evaluating the inhibitory activity of a test compound. In the CAM model, tumor cells invade through the chorioallantoic membrane containing CAM (with tumor cells in the presence of several serine protease inhibitors results in less or no invasion of the tumor cells through the membrane). Thus, the CAM assay is performed with CAM and tumor cells in the presence and absence of various concentrations of test compound. The invasiveness of tumor cells is measured under such conditions to provide an indication of the compound's inhibitory activity. A compound having inhibitory activity correlates with less tumor invasion.

protease inhibitor FUT-0175 and thromboxane synthetase inhibitor OKY-046 (Kaminogo *et al. Neurol. Med. Chir. (Tokyo) 38*:704-8; discussion 708-9 (1998)); The rat serine protease inhibitor 2.1 gene (LeCam, A., et al., *Biochem. Biophys. Res. Commun., 253*:311-4 (1998)); A new intracellular serine protease inhibitor expressed in the rat pituitary gland complexes with granzyme B (Hill *et al. FEBS Lett. 440*:361-4 (1998)); 3,4-Dichloroisocoumarin (Hammed *et al. Proc. Soc. Exp. Biol. Med., 219*:132-7 (1998)); LEX032 (Bains *et al. Eur. J. Pharmacol. 356*:67-72 (1998)); N-tosyl-L-phenylalanine chloromethyl ketone (Dryjanski *et al. Biochemistry 37*:14151-6 (1998)); Mouse gene for the serine protease inhibitor neuroserpin (P112) (Berger *et al. Gene, 214*:25-33 (1998)); Rat serine protease inhibitor 2.3 gene (Paul *et al. Eur. J. Biochem. 254*:538-46 (1998)); Ecotin (Yang *et al. J. Mol. Biol. 279*:945-57 (1998)); A 14 kDa plant-related serine protease inhibitor (Roch *et al. Dev. Comp. Immunol. 22*(1):1-12 (1998)); Matrix-associated serine protease inhibitor TFPI-2/33 kDa MSPI (Rao *et al. Int. J. Cancer 76*:749-56 (1998)); ONO-3403 (Hiwasa *et al. Cancer Lett. 126*:221-5 (1998)); Bdellastasin (Moser *et al. Eur. J. Biochem. 253*:212-20 (1998)); Bikunin (Xu *et al. J. Mol. Biol. 276*:955-66 (1998)); Nafamostat mesilate (Mellgren *et al. Thromb. Haemost. 79*:342-7 (1998)); The growth hormone dependent serine protease inhibitor, Spi 2.1 (Maake *et al. Endocrinology 138*:5630-6 (1997)); Growth factor activator inhibitor type 2, a Kunitz-type serine protease inhibitor (Kawaguchi *et al. J. Biol. Chem., 272*:27558-64 (1997)); Heat-stable serine protease inhibitor protein from ovaries of the desert locust, Schistocerga gregaria (Hamdaoui *et al. Biochem. Biophys. Res. Commun. 238*:357-60 (1997)); Human placental Hepatocyte growth factor activator inhibitor, a Kunitz-type serine protease inhibitor (Shimomura *et al. J. Biol. Chem. 272*:6370-6 (1997)); FUT-187, oral serine protease inhibitor (Shiozaki *et al. Gan To Kaguku Ryoho*, 23(14): 1971-9 (1996)); Extracellular matrix-associated serine protease inhibitors (Mr 33,000, 31,000, and 27,000 (Rao, C.N., et al., *Arch. Biochem. Biophys., 335*:82-92 (1996)); An irreversible isocoumarin serine protease inhibitor (Palencia, D.D., et al., *Biol. Reprod., 55*:536-42 (1996)); 4-(2-aminoethyl)-benzenesulfonyl fluoride (AEBSF) (Nakabo *et al. J. Leukoc. Biol. 60*:328-36 (1996)); Neuroserpin (Osterwalder, T., et al.,

P., *J. Immunol.*, 143(1):168-73 (1989)); Alpha 1-antichymotrypsin (Abraham, C.R., et al., *Cell*, 52(4):487-501 (1988)); Contrapsin (Modha, J., et al., *Parasitology*, 96 (Pt 1):99-109 (1988)); Alpha 2-antiplasmin (Holmes, W.E., et al., *J. Biol. Chem.*, 262(4):1659-64 (1987)); 3,4-dichloroisocoumarin (Harper, J.W., et al., *Biochemistry*, 24(8):1831-41 (1985)); Diisoprophylfluorophosphate (Tsutsui, K., et al., *Biochem. Biophys. Res. Commun.*, 123(1):271-7 (1984)); Gabexate mesilate (Hesse, B., et al., *Pharmacol. Res. Commun.*, 16(7):637-45 (1984)); Phenyl methyl sulfonyl fluoride (Dufer, J., et al., *Scand. J. Haematol.*, 32(1):25-32 (1984)); Protease inhibitor CI-2 (McPhalen, C.A., et al., *J. Mol. Biol.*, 168(2):445-7 (1983)); Phenylmethylsulfonyl fluoride (Sekar V., et al., *Biochem. Biophys. Res. Commun.*, 89(2):474-8 (1979)); PGE1 (Feinstein, M.D., et al., *Prostaglandine*, 14(6):1075-93 (1977).

**c. Combinatorial libraries and other libraries**

The source of compounds for the screening assays, can be libraries, including, but are not limited to, combinatorial libraries. Methods for synthesizing combinatorial libraries and characteristics of such combinatorial libraries are known in the art (*See generally, Combinatorial Libraries: Synthesis, Screening and Application Potential* (Cortese Ed.) Walter de Gruyter, Inc., 1995; Tietze and Lieb, *Curr. Opin. Chem. Biol.*, 2(3):363-71 (1998); Lam, *Anticancer Drug Des.*, 12(3):145-67 (1997); Blaney and Martin, *Curr. Opin. Chem. Biol.*, 1(1):54-9 (1997); and Schultz and Schultz, *Biotechnol. Prog.*, 12(6):729-43 (1996)).

Methods and strategies for generating diverse libraries, primarily peptide- and nucleotide-based oligomer libraries, have been developed using molecular biology methods and/or simultaneous chemical synthesis methodologies (*see, e.g.*, Dower et al., *Annu. Rep. Med. Chem.*, 26:271-280 (1991); Fodor et al., *Science*, 251:767-773 (1991); Jung et al., *Angew. Chem. Ind. Ed. Engl.*, 31:367-383 (1992); Zuckerman et al., *Proc. Natl. Acad. Sci. USA*, 89:4505-4509 (1992); Scott et al., *Science*, 249:386-390 (1990); Devlin et al., *Science*, 249:404-406 (1990); Cwirla et al., *Proc. Natl. Acad. Sci. USA*, 87:6378-6382 (1990); and Gallop et al., *J. Medicinal Chemistry*, 37:1233-1251 (1994)). The

2(12):29 (1996); Boublik, et al., Eukaryotic Virus Display: Engineering the Major Surface Glycoproteins of the Autographa California Nuclear Polyhedrosis Virus (ACNPV) for the Presentation of Foreign Proteins on the Virus Surface, *Bio/Technology*, 13:1079-1084 (1995); Brenner, et al., Encoded Combinatorial Chemistry, *Proc. Natl. Acad Sci. U.S.A.*, 89:5381-5383 (1992); Caflisch, et al., Computational Combinatorial Chemistry for *De Novo* Ligand Design: Review and Assessment, *Perspect. Drug Discovery Des.*, 3:51-84 (1995); Cheng, et al., Sequence-Selective Peptide Binding with a Peptido-A,B-*trans*-steroidal Receptor Selected from an Encoded Combinatorial Library, *J. Am. Chem. Soc.*, 118:1813-1814 (1996); Chu, et al., Affinity Capillary Electrophoresis to Identify the Peptide in A Peptide Library that Binds Most Tightly to Vancomycin, *J. Org. Chem.*, 58:648-652 (1993); Clackson, et al., Making Antibody Fragments Using Phage Display Libraries, *Nature*, 352:624-628 (1991); Combs, et al., Protein Structure-Based Combinatorial Chemistry: Discovery of Non-Peptide Binding Elements to Src SH3 Domain, *J. Am. Chem. Soc.*, 118:287-288 (1996); Cwirla, et al., Peptides On Phage: A Vast Library of Peptides for Identifying Ligands, *Proc. Natl. Acad. Sci. U.S.A.*, 87:6378-6382 (1990); Ecker, et al., Combinatorial Drug Discovery: Which Method will Produce the Greatest Value, *Bio/Technology*, 13:351-360 (1995); Ellington, et al., *In Vitro* Selection of RNA Molecules That Bind Specific Ligands, *Nature*, 346:818-822 (1990); Ellman, J.A., Variants of Benzodiazephines, *J. Am. Chem. Soc.*, 114:10997 (1992); Erickson, et al., *The Proteins*; Neurath, H., Hill, R.L., Eds.: Academic: New York, 1976; pp. 255-257; Felici, et al., *J. Mol. Biol.*, 222:301-310 (1991); Fodor, et al., Light-Directed, Spatially Addressable Parallel Chemical Synthesis, *Science*, 251:767-773 (1991); Francisco, et al., Transport and Anchoring of Beta-Lactamase to the External Surface of *E. Coli.*, *Proc. Natl. Acad. Sci. U.S.A.*, 89:2713-2717 (1992); Georgiou, et al., Practical Applications of Engineering Gram-Negative Bacterial Cell Surfaces, *TIBTECH*, 11:6-10 (1993); Geysen, et al., Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid, *Proc. Natl. Acad. Sci. U.S.A.*, 81:3998-4002 (1984); Glaser, et al., Antibody Engineering by Condon-Based Mutagenesis in a Filamentous Phage Vector System, *J. Immunol.*, 149:3903-3913 (1992); Gram, et al., In

of a Polypeptide Hormone, *Science*, 270:1657-1660 (1995); Light, et al., Display of Dimeric Bacterial Alkaline Phosphatase on the Major Coat Protein of Filamentous Bacteriophage, *Bioorg. Med. Chem. Lett.*, 3:1073-1079 (1992); Little, et al., Bacterial Surface Presentation of Proteins and Peptides: An Alternative to Phage Technology, *Trends Biotechnol.*, 11:3-5 (1993); Marks, et al., By-Passing Immunization. Human Antibodies from V-Gene Libraries Displayed on Phage, *J. Mol. Biol.*, 222:581-597 (1991); Matthews, et al., Substrate Phage: Selection of Protease Substrates by Monovalent Phage Display, *Science*, 260:1113-1117 (1993); McCafferty, et al., Phage Enzymes: Expression and Affinity Chromatography of Functional Alkaline Phosphatase on the Surface of Bacteriophage, *Protein Eng.*, 4:955-961 (1991); Menger, et al., Phosphatase Catalysis Developed Via Combinatorial Organic Chemistry, *J. Org. Chem.*, 60:6666-6667 (1995); Nicolaou, et al., *Angew. Chem. Int. Ed. Engl.*, 34:2289-2291 (1995); Oldenburg, et al., Peptide Ligands for A Sugar-Binding Protein Isolated from a Random Peptide Library, *Proc. Natl. Acad. Sci. U.S.A.*, 89:5393-5397 (1992); Parmley, et al., Antibody-Selectable Filamentous fd Phage Vectors: Affinity Purification of Target Genes, *Genes*, 73:305-318 (1988); Pinilla, et al., Synthetic Peptide Combinatorial Libraries (SPCLS)--Identification of the Antigenic Determinant of Beta-Endorphin Recognized by Monoclonal Antibody-3E7, *Gene*, 128:71-76 (1993); Pinilla, et al., Review of the Utility of Soluble Combinatorial Libraries, *Biopolymers*, 37:221-240 (1995); Pistor, et al., Expression of Viral Hemegglutinan On the Surface of *E. Coli.*, *Klin. Wochenschr.*, 66:110-116 (1989); Pollack, et al., Selective Chemical Catalysis by an Antibody, *Science*, 234:1570-1572 (1986); Rigler, et al., Fluorescence Correlations, Single Molecule Detection and Large Number Screening: Applications in Biotechnology, *J. Biotechnol.*, 41:177-186 (1995); Sarvetnick, et al., Increasing the Chemical Potential of the Germ-Line Antibody Repertoire, *Proc. Natl. Acad. Sci. U.S.A.*, 90:4008-4011 (1993); Sastry, et al., Cloning of the Immunological Repertiore in *Escherichia Coli* for Generation of Monoclonal Catalytic Antibodies: Construction of a Heavy Chain Variable Region-Specific cDNA Library, *Proc. Natl. Acad. Sci. U.S.A.*, 86:5728-5732 (1989); Scott, et al., Searching for Peptide Ligands with an Epitope Library, *Science*, 249:386-390 (1990); Sears, et al., Engineering

expression of the MTSP9 polypeptide, antibodies, peptide mimetics and other such compounds.

**1. Antibodies**

Antibodies, including polyclonal and monoclonal antibodies, that specifically bind to the MTSP9 polypeptide provided herein, particularly to the single chain protease domains thereof or the activated forms of the full-length or protease domain or the zymogen form, are provided.

Generally, the antibody is a monoclonal antibody, and typically the antibody specifically binds to the protease domain of the MTSP9 polypeptide. In particular embodiments, antibodies to each of the single chain and or two chain form of the protease domain of MTSP9 are provided. Also provided are antibodies that specifically bind to any domain of MTSP9 and to two chain forms thereof.

The MTSP9 polypeptide and domains, fragments, homologs and derivatives thereof can be used as immunogens to generate antibodies that specifically bind such immunogens. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library. In a specific embodiment, antibodies to human MTSP9 polypeptide are produced. In another embodiment, complexes formed from fragments of MTSP9 polypeptide, which fragments contain the serine protease domain, are used as immunogens for antibody production.

Various procedures known in the art can be used for the production of polyclonal antibodies to MTSP9 polypeptide, its domains, derivatives, fragments or analogs. For production of the antibody, various host animals can be immunized by injection with the native MTSP9 polypeptide or a synthetic version, or a derivative of the foregoing, such as a cross-linked MTSP9 polypeptide. Such host animals include but are not limited to rabbits, mice, rats, etc. Various adjuvants can be used to increase the immunological response, depending on the host species, and include but are not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil

thereof. Non-human antibodies can be "humanized" by known methods (*see, e.g.*, U.S. Patent No. 5,225,539).

Antibody fragments that specifically bind to MTSP9 polyeptide or epitopes thereof can be generated by techniques known in the art. For example, such fragments include but are not limited to: the F(ab')2 fragment, which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments that can be generated by reducing the disulfide bridges of the F(ab')2 fragment, the Fab fragments that can be generated by treating the antibody molecular with papain and a reducing agent, and Fv fragments.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, *e.g.*, ELISA (enzyme-linked immunosorbent assay). To select antibodies specific for a particular domain of the MTSP9 polypeptide one can assay generated hybridomas for a product that binds to the fragment of the MTSP9 polypeptide that contains such a domain.

The foregoing antibodies can be used in methods known in the art relating to the localization and/or quantitation of MTSP9 polypeptide proteins, *e.g.*, for imaging these proteins, measuring levels thereof in appropriate physiological samples, in, for example, diagnostic methods. In another embodiment, anti-MTSP9 polypeptide antibodies, or fragments thereof, containing the binding domain are used as therapeutic agents.

**2. Peptides, Polypeptides and Peptide Mimetics**

Provided herein are methods for identifying molecules that bind to and modulate the activity of SP proteins. Included among molecules that bind to SPs, particularly the single chain protease domain or catalytically active fragments thereof, are peptides, polypeptides and peptide mimetics, including cyclic peptides. Peptide mimetics are molecules or compounds that mimic the necessary molecular conformation of a ligand or polypeptide for specific binding to a target molecule such as an MTSP9 polypeptide. In an exemplary embodiment, the peptides, peptides, polypeptides and peptide mimetics or peptide mimetics bind to the protease domain of the MTSP9 polypeptide. Such peptides and peptide mimetics include those of antibodies that specifically bind to an MTSP9 polypeptide and, typically, bind to the protease domain of an

on the patient's state of health and weight. Accordingly, the peptides, polypeptides and peptide mimetics that bind to an MTSP9 polypeptide can be used to prepare pharmaceutical compositions containing, as an active ingredient, at least one of the peptides or peptide mimetics in association with a pharmaceutical carrier or diluent. The compounds can be administered, for example, by oral, pulmonary, parental (intramuscular, intraperitoneal, intravenous (IV) or subcutaneous injection), inhalation (via a fine powder formulation), transdermal, nasal, vaginal, rectal, or sublingual routes of administration and can be formulated in dosage forms appropriate for each route of administration (see, *e.g.*, International PCT application Nos. WO 93/25221 and WO 94/17784; and European Patent Application 613,683).

Peptides, polypeptides and peptide mimetics that bind to MTSP9 polypeptides are useful *in vitro* as unique tools for understanding the biological role of MTSP9 polypeptides, including the evaluation of the many factors thought to influence, and be influenced by, the production of MTSP9 polypeptide. Such peptides, polypeptides and peptide mimetics are also useful in the development of other compounds that bind to and modulate the activity of an MTSP9 polypeptide, because such compounds provide important information on the relationship between structure and activity that should facilitate such development.

The peptides, polypeptides and peptide mimetics are also useful as competitive binders in assays to screen for new MTSP9 polypeptides or MTSP9 polypeptide agonists. In such assay embodiments, the compounds can be used without modification or can be modified in a variety of ways; for example, by labeling, such as covalently or non-covalently joining a moiety which directly or indirectly provides a detectable signal. In any of these assays, the materials thereto can be labeled either directly or indirectly. Possibilities for direct labeling include label groups such as: radiolabels such as $^{125}I$ enzymes (U.S. Pat. No. 3,645,090) such as peroxidase and alkaline phosphatase, and fluorescent labels (U.S. Pat. No. 3,940,475) capable of monitoring the change in fluorescence intensity, wavelength shift, or fluorescence polarization. Possibilities for indirect labeling include biotinylation of one constituent followed by binding to avidin

equivalent or enhanced therapeutic or prophylactic effect. Preparation of peptidomimetics and structures thereof are known to those of skill in this art.

Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) can be used to generate more stable peptides. In addition, constrained peptides containing a consensus sequence or a substantially identical consensus sequence variation can be generated by methods known in the art (Rizo *et al.* (1992) *An. Rev. Biochem., 61*:387, incorporated herein by reference); for example, by adding internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

Those skilled in the art appreciate that modifications can be made to the peptides and mimetics without deleteriously effecting the biological or functional activity of the peptide. Further, the skilled artisan would know how to design non-peptide structures in three dimensional terms, that mimic the peptides that bind to a target molecule, *e.g.*, an MTSP9 polypeptide or, generally, the protease domain of MTSP9 polypeptides (see, *e.g.*, Eck and Sprang (1989) *J. Biol. Chem., 26*: 17605-18795).

When used for diagnostic purposes, the peptides and peptide mimetics can be labeled with a detectable label and, accordingly, the peptides and peptide mimetics without such a label can serve as intermediates in the preparation of labeled peptides and peptide mimetics. Detectable labels can be molecules or compounds, which when covalently attached to the peptides and peptide mimetics, permit detection of the peptide and peptide mimetics *in vivo*, for example, in a patient to whom the peptide or peptide mimetic has been administered, or *in vitro*, *e.g.*, in a sample or cells. Suitable detectable labels are well known in the art and include, by way of example, radioisotopes, fluorescent labels (*e.g.*, fluorescein), and the like. The particular detectable label employed is not critical and is selected to be detectable at non-toxic levels. Selection of the such labels is well within the skill of the art.

Covalent attachment of a detectable label to the peptide or peptide mimetic is accomplished by conventional methods well known in the art. For example, when the $^{125}I$ radioisotope is employed as the detectable label, covalent

The peptides or peptide mimetics can be introduced into cells, *in vivo* or *ex vivo*, by microinjection or by use of liposomes, for example. Alternatively, the peptides, polypeptides or peptide mimetics can be taken up by cells, *in vivo* or *ex vivo*, actively or by diffusion. In addition, extracellular application of the peptide, polypeptide or peptide mimetic can be sufficient to effect treatment of a neoplastic disease. Other molecules, such as drugs or organic compounds, that: 1) bind to a MTSP9 polypeptide or protease domain thereof; or 2) have a similar function or activity to an MTSP9 polypeptide or protease domain thereof, can be used in methods for treatment.

**4. Rational drug design**

The goal of rational drug design is to produce structural analogs of biologically active polypeptides or peptides of interest or of small molecules or peptide mimetics with which they interact (*e.g.*, agonists and antagonists) in order to fashion drugs which are, e.g., more active or stable forms thereof; or which, for example, enhance or interfere with the function of a polypeptide *in vivo* (e.g., an MTSP9 polypeptide). In one approach, one first determines the three-dimensional structure of a protein of interest (e.g., an MTSP9 polypeptide or polypeptide having a protease domain) or, for example, of an MTSP9 polypeptide-ligand complex, by X-ray crystallography, by computer modeling or most typically, by a combination of approaches (see, *e.g.*, Erickson *et al.* 1990). Also, useful information regarding the structure of a polypeptide can be gained by modeling based on the structure of homologous proteins. In addition, peptides can be analyzed by an alanine scan. In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

Also, a polypeptide or peptide that binds to an MTSP9 polypeptide or, generally, the protease domain of an MTSP9 polypeptide, can be selected by a functional assay, and then the crystal structure of this polypeptide or peptide can be determined. The polypeptide can be, for example, an antibody specific for an MTSP9 polypeptide or the protein domain of an MTSP9 polypeptide. This approach can yield a pharmacophore upon which subsequent drug design can be

large scale immobilized polymer synthesis" system (see, *e.g.*, U.S. Patent No. 6,121,238; and Dower *et al.* (1991) *An. Rep. Med. Chem. 26*:271-280

For example, using the procedures described above, random peptides can generally be designed to have a defined number of amino acid residues in length (*e.g.*, 12). To generate the collection of oligonucleotides encoding the random peptides, the codon motif (NNK)x, where N is nucleotide A, C, G, or T (equimolar; depending on the methodology employed, other nucleotides can be employed), K is G or T (equimolar), and x is an integer corresponding to the number of amino acids in the peptide (*e.g.*, 12) can be used to specify any one of the 32 possible codons resulting from the NNK motif: 1 for each of 12 amino acids, 2 for each of 5 amino acids, 3 for each of 3 amino acids, and only one of the three stop codons. Thus, the NNK motif encodes all of the amino acids, encodes only one stop codon, and reduces codon bias.

The random peptides can be presented, for example, either on the surface of a phage particle, as part of a fusion protein containing either the pIII or the pVIII coat protein of a phage fd derivative (peptides on phage) or as a fusion protein with the LacI peptide fusion protein bound to a plasmid (peptides on plasmids). The phage or plasmids, including the DNA encoding the peptides, can be identified and isolated by an affinity enrichment process using immobilized MTSP9 polypeptide having a protease domain. The affinity enrichment process, sometimes called "panning," typically involves multiple rounds of incubating the phage, plasmids, or polysomes with the immobilized MTSP9 polypeptide, collecting the phage, plasmids, or polysomes that bind to the MTSP9 polypeptide (along with the accompanying DNA or mRNA), and producing more of the phage or plasmids (along with the accompanying LacI-peptide fusion protein) collected.

**Characteristics of peptides and peptide mimetics**

Among the peptides, polypeptides and peptide mimetics for therapeutic application are those of having molecular weights from about 250 to about 8,000 daltons. If such peptides are oligomerized, dimerized and/or derivatized with a hydrophilic polymer (*e.g.*, to increase the affinity and/or activity of the compounds), the molecular weights of such peptides can be substantially greater and can range anywhere from about 500 to about 120,000 daltons, generally

synthesize truncation analogs and deletion analogs and combinations of truncation and deletion analogs of the peptide compounds.

These procedures can also be used to synthesize peptides in which amino acids other than the 20 naturally occurring, genetically encoded amino acids are substituted at one, two, or more positions of the peptide. For instance, naphthylalanine can be substituted for tryptophan, facilitating synthesis. Other synthetic amino acids that can be substituted into the peptides include L-hydroxypropyl, L-3, 4-dihydroxy-phenylalanyl, d amino acids such as L-d-hydroxylysyl and D-d-methylalanyl, L-$\alpha$-methylalanyl, $\beta$ amino acids, and isoquinolyl. D amino acids and non-naturally occurring synthetic amino acids can also be incorporated into the peptides (see, *e.g.*, Roberts *et al.* (1983) *Unusual Amino/Acids in Peptide Synthesis, 5*(6):341-449).

The peptides also can be modified by phosphorylation (see, *e.g.*, W. Bannwarth *et al.* (1996) *Biorganic and Medicinal Chemistry Letters, 6*(17):2141-2146), and other methods for making peptide derivatives (see, *e.g.*, Hruby *et al.* (1990) *Biochem. J., 268*(2):249-262). Thus, peptide compounds also serve as a basis to prepare peptide mimetics with similar or improved biological activity.

Those of skill in the art recognize that a variety of techniques are available for constructing peptide mimetics with the same or similar desired biological activity as the corresponding peptide compound but with more favorable activity than the peptide with respect to solubility, stability, and susceptibility to hydrolysis and proteolysis (see, *e.g.*, Morgan *et al.* (1989) *An. Rep. Med. Chem., 24*:243-252). Methods for preparing peptide mimetics modified at the N-terminal amino group, the C-terminal carboxyl group, and/or changing one or more of the amido linkages in the peptide to a non-amido linkage are known to those of skill in the art.

Amino terminus modifications include, but are not limited to, alkylating, acetylating and adding a carbobenzoyl group, forming a succinimide group (see, *e.g.*, Murray *et al.* (1995) *Burger's Medicinal Chemistry and Drug Discovery, 5th ed., Vol. 1,* Manfred E. Wolf, ed., John Wiley and Sons, Inc.). C-terminal

reference). These techniques include replacing the peptide backbone with a backbone composed of phosphonates, amidates, carbamates, sulfonamides, secondary amines, and N-methylamino acids.

Peptide compounds can exist in a cyclized form with an intramolecular disulfide bond between the thiol groups of the cysteines. Alternatively, an intermolecular disulfide bond between the thiol groups of the cysteines can be produced to yield a dimeric (or higher oligomeric) compound. One or more of the cysteine residues can also be substituted with a homocysteine.

**I. Conjugates**

A conjugate, containing: a) a single chain protease domain (or proteolytically active portion thereof) of an MTSP9 polypeptide or a full length zymogen, activated form thereof, or two or single chain protease domain thereof; and b) a targeting agent linked to the MTSP9 polypeptide directly or via a linker, wherein the agent facilitates: i) affinity isolation or purification of the conjugate; ii) attachment of the conjugate to a surface; iii) detection of the conjugate; or iv) targeted delivery to a selected tissue or cell, is provided herein. The conjugate can be a chemical conjugate or a fusion protein mixture thereof.

The targeting agent can be a protein or peptide fragment, such as a tissue specific or tumor specific monoclonal antibody or growth factor or fragment thereof linked either directly or via a linker to an MTSP9 polypeptide or a protease domain thereof. The targeting agent can also be a protein or peptide fragment that contains a protein binding sequence, a nucleic acid binding sequence, a lipid binding sequence, a polysaccharide binding sequence, or a metal binding sequence, or a linker for attachment to a solid support. In a particular embodiment, the conjugate contains a) the MTSP9 or portion thereof, as described herein; and b) a targeting agent linked to the MTSP9 polypeptide directly or via a linker.

Conjugates, such as fusion proteins and chemical conjugates, of the MTSP9 polypeptide with a protein or peptide fragment (or plurality thereof) that functions, for example, to facilitate affinity isolation or purification of the MTSP9 polypeptide domain, attachment of the MTSP9 polypeptide domain to a surface, or detection of the MTSP9 polypeptide domain are provided. The conjugates can

This page is not part of

the pamphlet!

# WO 02-077267

# 2/2

Date: 03 oct 2002

Destination: Agent

Address:

a. **Fusion proteins**

Fusion proteins are provided herein. A fusion protein contains: a) one or a plurality of domains of an MTSP9 polypeptide and b) a targeting agent. The fusion proteins are generally produced by recombinant expression of nucleic acids that encode the fusion protein.

b. **Chemical conjugation**

To effect chemical conjugation herein, the MTSP9 polypeptide domain is linked via one or more selected linkers or directly to the targeting agent. Chemical conjugation must be used if the targeted agent is other than a peptide or protein, such as a nucleic acid or a non-peptide drug. Any means known to those of skill in the art for chemically conjugating selected moieties can be used.

2. **Linkers**

Linkers for two purposes are contemplated herein. The conjugates can include one or more linkers between the MTSP9 polypeptide portion and the targeting agent. Additionally, linkers are used for facilitating or enhancing immobilization of an MTSP9 polypeptide or portion thereof on a solid support, such as a microtiter plate, silicon or silicon-coated chip, glass or plastic support, such as for high throughput solid phase screening protocols.

Any linker known to those of skill in the art for preparation of conjugates can be used herein. These linkers are typically used in the preparation of chemical conjugates; peptide linkers can be incorporated into fusion proteins.

Linkers can be any moiety suitable to associate a domain of MTSP9 polypeptide and a targeting agent. Such linkers and linkages include, but are not limited to, peptidic linkages, amino acid and peptide linkages, typically containing between one and about 60 amino acids, more generally between about 10 and 30 amino acids, chemical linkers, such as heterobifunctional cleavable cross-linkers, including but are not limited to, N-succinimidyl (4-iodoacetyl)-aminobenzoate, sulfosuccinimydil (4-iodoacetyl)-aminobenzoate, 4-succinimidyl-oxycarbonyl-a- (2-pyridyldithio)toluene, sulfosuccinimidyl-6- [a-methyl-a-(pyridyldithiol)-toluamido] hexanoate, N-succinimidyl-3-(-2-pyridyldithio) -proprionate, succinimidyl 6[3(-(-2-pyridyldithio)-proprionamido] hexanoate, sulfosuccinimidyl 6[3(-(-2-pyridyldithio)-propionamido] hexanoate, 3-(2-pyridyldi-

**a) Acid cleavable, photocleavable and heat sensitive linkers**

Acid cleavable linkers, photocleavable and heat sensitive linkers can also be used, particularly where it can be necessary to cleave the domain of MTSP9 polypeptide to permit it to be more readily accessible to reaction. Acid cleavable linkers include, but are not limited to, bismaleimideothoxy propane; and adipic acid dihydrazide linkers (see, *e.g.*, Fattom *et al.* (1992) *Infection & Immun. 60*:584-589) and acid labile transferrin conjugates that contain a sufficient portion of transferrin to permit entry into the intracellular transferrin cycling pathway (see, *e.g.*, Welhöner *et al.* (1991) *J. Biol. Chem. 266*:4309-4314).

Photocleavable linkers are linkers that are cleaved upon exposure to light (see, *e.g.*, Goldmacher *et al.* (1992) *Bioconj. Chem. 3*:104-107, which linkers are herein incorporated by reference), thereby releasing the targeted agent upon exposure to light. Photocleavable linkers that are cleaved upon exposure to light are known (see, *e.g.*, Hazum *et al.* (1981) in *Pept., Proc. Eur. Pept. Symp.*, 16th, Brunfeldt, K (Ed), pp. 105-110, which describes the use of a nitrobenzyl group as a photocleavable protective group for cysteine; Yen *et al.* (1989) *Makromol. Chem 190*:69-82, which describes water soluble photocleavable copolymers, including hydroxypropylmethacrylamide copolymer, glycine copolymer, fluorescein copolymer and methylrhodamine copolymer; Goldmacher *et al.* (1992) *Bioconj. Chem. 3*:104-107, which describes a cross-linker and reagent that undergoes photolytic degradation upon exposure to near UV light (350 nm); and Senter *et al.* (1985) *Photochem. Photobiol 42*:231-237, which describes nitrobenzyloxycarbonyl chloride cross linking reagents that produce photocleavable linkages), thereby releasing the targeted agent upon exposure to light. Such linkers would have particular use in treating dermatological or ophthalmic conditions that can be exposed to light using fiber optics. After administration of the conjugate, the eye or skin or other body part can be exposed to light, resulting in release of the targeted moiety from the conjugate. Such photocleavable linkers are useful in connection with diagnostic

3. **Targeting agents**

Any agent that facilitates detection, immobilization, or purification of the conjugate is contemplated for use herein. For chemical conjugates any moiety that has such properties is contemplated; for fusion proteins, the targeting agent is a protein, peptide or fragment thereof that is sufficient to effects the targeting activity. Contemplated targeting agents include those that deliver the MTSP9 polypeptide or portion thereof to selected cells and tissues. Such agents include tumor specific monoclonal antibodies and portions thereof, growth factors, such as FGF, EGF, PDGF, VEGF, cytokines, including chemokines, and other such agents.

4. **Nucleic acids, plasmids and cells**

Isolated nucleic acid fragments encoding fusion proteins are provided. The nucleic acid fragment that encodes the fusion protein includes: a) nucleic acid encoding a protease domain of an MTSP9 polypeptide; and b) nucleic acid encoding a protein, peptide or effective fragment thereof that facilitates: i) affinity isolation or purification of the fusion protein; ii) attachment of the fusion protein to a surface; or iii) detection of the fusion protein. Generally, the nucleic acid is DNA.

Plasmids for replication and vectors for expression that contain the above nucleic acid fragments are also provided. Cells containing the plasmids and vectors are also provided. The cells can be any suitable host including, but are not limited to, bacterial cells, yeast cells, fungal cells, plant cells, insect cell and animal cells. The nucleic acids, plasmids, and cells containing the plasmids can be prepared according to methods known in the art including any described herein.

Also provided are methods for producing the above fusion proteins. An exemplary method includes the steps of growing, for example, culturing the cells so that they proliferate, cells containing a plasmid encoding the fusion protein under conditions whereby the fusion protein is expressed by the cell, and recovering the expressed fusion protein. Methods for expressing and recovering recombinant proteins are well known in the art (*See generally, Current Protocols*

syntheses and separations. Such supports are used, for example, in affinity chromatography, in the immobilization of biologically active materials, and during chemical syntheses of biomolecules, including proteins, amino acids and other organic molecules and polymers. The preparation of and use of supports is well known to those of skill in this art; there are many such materials and preparations thereof known. For example, naturally-occurring support materials, such as agarose and cellulose, can be isolated from their respective sources, and processed according to known protocols, and synthetic materials can be prepared in accord with known protocols.

The supports are typically insoluble materials that are solid, porous, deformable, or hard, and have any required structure and geometry, including, but not limited to: beads, pellets, disks, capillaries, hollow fibers, needles, solid fibers, random shapes, thin films and membranes. Thus, the item can be fabricated from the matrix material or combined with it, such as by coating all or part of the surface or impregnating particles.

Typically, when the matrix is particulate, the particles are at least about 10-2000 $\mu$m, but can be smaller or larger, depending upon the selected application. Selection of the matrices is governed, at least in part, by their physical and chemical properties, such as solubility, functional groups, mechanical stability, surface area swelling propensity, hydrophobic or hydrophilic properties and intended use.

If necessary, the support matrix material can be treated to contain an appropriate reactive moiety. In some cases, the support matrix material already containing the reactive moiety can be obtained commercially. The support matrix material containing the reactive moiety can thereby serve as the matrix support upon which molecules are linked. Materials containing reactive surface moieties such as amino silane linkages, hydroxyl linkages or carboxysilane linkages can be produced by well established surface chemistry techniques involving silanization reactions, or the like. Examples of these materials are those having surface silicon oxide moieties, covalently linked to gamma-amino-propylsilane, and other organic moieties; N-[3-(triethyoxysilyl)propyl]phthelamic acid; and bis-(2-hydroxyethyl)aminopropyltriethoxysilane. Exemplary of readily

polymers and co-polymers such as polyvinylalcohols, acrylates and acrylic acids such as polyethylene-co-acrylic acid, polyethylene-co-methacrylic acid, polyethylene-co-ethylacrylate, polyethylene-co-methyl acrylate, polypropylene-co-acrylic acid, polypropylene-co-methyl-acrylic acid, polypropylene-co-ethylacrylate, polypropylene-co-methyl acrylate, polyethylene-co-vinyl acetate, polypropylene-co-vinyl acetate, and those containing acid anhydride groups such as polyethylene-co-maleic anhydride and polypropylene-co-maleic anhydride. Liposomes have also been used as solid supports for affinity purifications (Powell et al. *Biotechnol. Bioeng.*, 33:173 (1989)).

Numerous methods have been developed for the immobilization of proteins and other biomolecules onto solid or liquid supports (*see, e.g.*, Mosbach, *Methods in Enzymology*, 44 (1976); Weetall, *Immobilized Enzymes, Antigens, Antibodies, and Peptides*, (1975); Kennedy et al., *Solid Phase Biochemistry, Analytical and Synthetic Aspects*, Scouten, ed., pp. 253-391 (1983); *see, generally*, Affinity Techniques. Enzyme Purification: *Part B. Methods in Enzymology*, Vol. 34, ed. W. B. Jakoby, M. Wilchek, Acad. Press, N.Y. (1974); and Immobilized Biochemicals and Affinity Chromatography, *Advances in Experimental Medicine and Biology*, vol. 42, ed. R. Dunlap, Plenum Press, N.Y. (1974)).

Among the most commonly used methods are absorption and adsorption or covalent binding to the support, either directly or via a linker, such as the numerous disulfide linkages, thioether bonds, hindered disulfide bonds, and covalent bonds between free reactive groups, such as amine and thiol groups, known to those of skill in art (*see, e.g.*, the PIERCE CATALOG, ImmunoTechnology Catalog & Handbook, 1992-1993, which describes the preparation of and use of such reagents and provides a commercial source for such reagents; Wong, *Chemistry of Protein Conjugation and Cross Linking*, CRC Press (1993); *see also* DeWitt et al., *Proc. Natl. Acad. Sci. U.S.A.*, 90:6909 (1993); Zuckermann et al., *J. Am. Chem. Soc.*, 114:10646 (1992); Kurth et al., *J. Am. Chem. Soc.*, 116:2661 (1994); Ellman et al., *Proc. Natl. Acad. Sci. U.S.A.*, 91:4708 (1994); Sucholeiki, *Tetrahedron Lttrs.*, 35:7307 (1994); Su-Sun Wang, *J. Org. Chem.*, 41:3258 (1976); Padwa et al., *J. Org. Chem.*,

reactions, gel diffusion precipitin reactions, immunodiffusion assays, agglutination assays, complement-fixation assays, immunoradiometric assays, fluorescent immunoassays and protein A immunoassays.

MTSP9 polypeptide genes and related nucleic acid sequences and subsequences, including complementary sequences, also can be used in hybridization assays. MTSP9 polypeptide nucleic acid sequences, or subsequences thereof containing about at least 8 nucleotides, generally 14 or 16 or 30 or more, generally less than 1000 or up to 100, continugous nucleotides can be used as hybridization probes. Hybridization assays can be used to detect, prognose, diagnose, or monitor conditions, disorders, or disease states associated with aberrant changes in MTSP9 polypeptide expression and/or activity as described herein. In particular, such a hybridization assay is carried out by a method by contacting a sample containing nucleic acid with a nucleic acid probe capable of hybridizing to MTSP9 polypeptide encoding DNA or RNA, under conditions such that hybridization can occur, and detecting or measuring any resulting hybridization.

In a specific embodiment, a method of diagnosing a disease or disorder characterized by detecting an aberrant level of an MTSP9 polypeptide in a subject is provided herein by measuring the level of the DNA, RNA, protein or functional activity of the MTSP9 polypeptide in a sample derived from the subject, wherein an increase or decrease in the level of the DNA, RNA, protein or functional activity of the MTSP9 polypeptide, relative to the level of the DNA, RNA, protein or functional activity found in an analogous sample not having the disease or disorder indicates the presence of the disease or disorder in the subject.

Kits for diagnostic use are also provided, that contain in one or more containers an anti-MTSP9 polypeptide antibody, and, optionally, a labeled binding partner to the antibody. Alternatively, the anti-MTSP9 polypeptide antibody can be labeled (with a detectable marker, *e.g.*, a chemiluminescent, enzymatic, fluorescent, or radioactive moiety). A kit is also provided that includes in one or more containers a nucleic acid probe capable of hybridizing to the MTSP9 polypeptide-encoding nucleic acid. In a specific embodiment, a kit

inhibitor of MTSP9 polypeptide membrane-localization, or any inhibitor of the expression of or, especially, the activity of an MTSP9 polypeptide.

**b. Anti-angiogenic agents and anti-tumor agents**

Any anti-angiogenic agents and anti-tumor agents, including those described herein, when used alone or in combination with other compounds, that can alleviate, reduce, ameliorate, prevent, or place or maintain in a state of remission of clinical symptoms or diagnostic markers associated with undesired and/or uncontrolled angiogenesis and/or tumor growth and metastasis, particularly solid neoplasms, vascular malformations and cardiovascular disorders, chronic inflammatory diseases and aberrant wound repairs, circulatory disorders, crest syndromes, dermatological disorders, or ocular disorders, can be used in the combinations. Also contemplated are anti-tumor agents for use in combination with an inhibitor of an MTSP9 polypeptide.

**c. Anti-tumor agents and anti-angiogenic agents**

The compounds identified by the methods provided herein or provided herein can be used in combination with anti-tumor agents and/or anti-angiogenesis agents.

**2. Formulations and route of administration**

The compounds herein and agents can be formulated as pharmaceutical compositions, typically for single dosage administration. The concentrations of the compounds in the formulations are effective for delivery of an amount, upon administration, that is effective for the intended treatment. Typically, the compositions are formulated for single dosage administration. To formulate a composition, the weight fraction of a compound or mixture thereof is dissolved, suspended, dispersed or otherwise mixed in a selected vehicle at an effective concentration such that the treated condition is relieved or ameliorated. Pharmaceutical carriers or vehicles suitable for administration of the compounds provided herein include any such carriers known to those skilled in the art to be suitable for the particular mode of administration.

In addition, the compounds can be formulated as the sole pharmaceutically active ingredient in the composition or can be combined with other active ingredients. Liposomal suspensions, including tissue-targeted

Pharmaceutically acceptable derivatives include acids, salts, esters, hydrates, solvates and prodrug forms. The derivative is typically selected such that its pharmacokinetic properties are superior to the corresponding neutral compound.

Thus, effective concentrations or amounts of one or more of the compounds provided herein or pharmaceutically acceptable derivatives thereof are mixed with a suitable pharmaceutical carrier or vehicle for systemic, topical or local administration to form pharmaceutical compositions. Compounds are included in an amount effective for ameliorating or treating the disorder for which treatment is contemplated. The concentration of active compound in the composition depends on absorption, inactivation, excretion rates of the active compound, the dosage schedule, amount administered, particular formulation as well as other factors known to those of skill in the art.

Solutions or suspensions used for parenteral, intradermal, subcutaneous, or topical application can include any of the following components: a sterile diluent, such as water for injection, saline solution, fixed oil, polyethylene glycol, glycerine, propylene glycol or other synthetic solvent; antimicrobial agents, such as benzyl alcohol and methyl parabens; antioxidants, such as ascorbic acid and sodium bisulfite; chelating agents, such as ethylenediaminetetraacetic acid (EDTA); buffers, such as acetates, citrates and phosphates; and agents for the adjustment of tonicity such as sodium chloride or dextrose. Parenteral preparations can be enclosed in ampules, disposable syringes or single or multiple dose vials made of glass, plastic or other suitable material.

In instances in which the compounds exhibit insufficient solubility, methods for solubilizing compounds can be used. Such methods are known to those of skill in this art, and include, but are not limited to, using cosolvents, such as dimethylsulfoxide (DMSO), using surfactants, such as Tween®, or dissolution in aqueous sodium bicarbonate. Derivatives of the compounds, such as prodrugs of the compounds can also be used in formulating effective pharmaceutical compositions. For ophthalmic indications, the compositions are formulated in an ophthalmically acceptable carrier. For the ophthalmic uses herein, local administration, either by topical administration or by injection are

for human and animal subjects and packaged individually as is known in the art. Each unit-dose contains a predetermined quantity of the therapeutically active compound sufficient to produce the desired therapeutic effect, in association with the required pharmaceutical carrier, vehicle or diluent. Examples of unit-dose forms include ampoules and syringes and individually packaged tablets or capsules. Unit-dose forms can be administered in fractions or multiples thereof. A multiple-dose form is a plurality of identical unit-dosage forms packaged in a single container to be administered in segregated unit-dose form. Examples of multiple-dose forms include vials, bottles of tablets or capsules or bottles of pints or gallons. Hence, multiple dose form is a multiple of unit-doses which are not segregated in packaging.

The composition can contain along with the active ingredient: a diluent such as lactose, sucrose, dicalcium phosphate, or carboxymethylcellulose; a lubricant, such as magnesium stearate, calcium stearate and talc; and a binder such as starch, natural gums, such as gum acaciagelatin, glucose, molasses, polvinylpyrrolidine, celluloses and derivatives thereof, povidone, crospovidones and other such binders known to those of skill in the art. Liquid pharmaceutically administrable compositions can, for example, be prepared by dissolving, dispersing, or otherwise mixing an active compound as defined above and optional pharmaceutical adjuvants in a carrier, such as, for example, water, saline, aqueous dextrose, glycerol, glycols, ethanol, and the like, to thereby form a solution or suspension. If desired, the pharmaceutical composition to be administered can also contain minor amounts of nontoxic auxiliary substances such as wetting agents, emulsifying agents, or solubilizing agents, pH buffering agents and the like, for example, acetate, sodium citrate, cyclodextrine derivatives, sorbitan monolaurate, triethanolamine sodium acetate, triethanolamine oleate, and other such agents. Methods of preparing such dosage forms are known, or will be apparent, to those skilled in this art (see, *e.g.*, Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., 15th Edition, 1975). The composition or formulation to be administered contains a quantity of the active compound in an amount sufficient to alleviate the symptoms of the treated subject.

formulation is often applied by instillation or as an ointment into the conjunctival sac. It also can be used for irrigation or lubrication of the eye, facial sinuses, and external auditory meatus. It can also be injected into the anterior eye chamber and other places. The topical formulations in the liquid state can be also present in a hydrophilic three-dimensional polymer matrix in the form of a strip, contact lens, and the like from which the active components are released.

For administration by inhalation, the compounds for use herein can be delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol, the dosage unit can be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin, for use in an inhaler or insufflator can be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

Formulations suitable for buccal (sublingual) administration include, for example, lozenges containing the active compound in a flavored base, usually sucrose and acacia or tragacanth; and pastilles containing the compound in an inert base such as gelatin and glycerin or sucrose and acacia.

The compounds can be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection can be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions can be suspensions, solutions or emulsions in oily or aqueous vehicles, and can contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient can be in powder form for reconstitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water or other solvents, before use.

Formulations suitable for transdermal administration can be presented as discrete patches adapted to remain in intimate contact with the epidermis of the recipient for a prolonged period of time. Such patches suitably contain the active compound as an optionally buffered aqueous solution of, for example, 0.1 to 0.2 M concentration with respect to the active compound. Formulations

diseases or disorders contemplated herein. The label can optionally include the disorders for which the therapy is warranted.

**L. Methods of treatment**

The compounds identified by the methods herein are used for treating or preventing neoplastic diseases in an animal, particularly a mammal, including a human, is provided herein. In one embodiment, the method includes administering to a mammal an effective amount of an inhibitor of an MTSP9 polypeptide, whereby the disease or disorder is treated or prevented.

In an embodiment, the MTSP9 polypeptide inhibitor used in the treatment or prevention is administered with a pharmaceutically acceptable carrier or excipient. The mammal treated can be a human. The inhibitors provided herein are those identified by the screening assays. In addition, antibodies and antisense nucleic acids or double-stranded RNA (dsRNA), such as RNAi, are contemplated.

The treatment or prevention method can further include administering an anti-angiogenic treatment or agent or anti-tumor agent simultaneously with, prior to or subsequent to the MTSP9 polypeptide inhibitor, which can be any compound identified that inhibits the activity of an MTSP9 polypeptide. Such compounds include small molecule modulators, an antibody or a fragment or derivative thereof containing a binding region thereof against the MTSP9 polypeptide, an antisense nucleic acid or double-stranded RNA (dsRNA), such as RNAi, encoding an a portion of the MTSP9 polypeptide or complementary to thereto, and a nucleic acid containing at least a portion of a gene encoding the MTSP9 polypeptide into which a heterologous nucleotide sequence has been inserted such that the heterologous sequence inactivates the biological activity of at least a portion of the gene encoding the MTSP9 polypeptide, in which the portion of the gene encoding the MTSP9 polypeptide flanks the heterologous sequence to promote homologous recombination with a genomic gene encoding the MTSP9 polypeptide. In addition, such molecules are generally less than about 1000 nt long.

The MTSP9 polypeptide antisense nucleic acid generally is an oligonucleotide, typically single-stranded DNA or RNA or an analog thereof or mixtures thereof. For example, the oligonucleotide includes a sequence antisense to a portion of a nucleic acid that encodes a human MTSP9 polypeptide. The oligonucleotide can be modified at any position on its structure with substituents generally known in the art.

The MTSP9 polypeptide antisense oligonucleotide can include at least one modified base moiety which is selected from the group including, but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

In another embodiment, the oligonucleotide includes at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose. The oligonucleotide can include at least one modified phosphate backbone selected from a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

The oligonucleotide can be an $\alpha$-anomeric oligonucleotide. An $\alpha$-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which the strands run parallel to each other (Gautier et al., *Nucl. Acids Res.* 15:6625-6641 (1987)).

(Bernoist and Chambon, *Nature* 290:304-310 (1981), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto et al., *Cell* 22:787-797 (1980), the herpes thymidine kinase promoter (Wagner et al., *Proc. Natl. Acad. Sci. U.S.A.* 78:1441-1445 (1981), the regulatory sequences of the metallothionein gene (Brinster et al., *Nature* 296:39-42 (1982), etc.

The antisense nucleic acids include sequence complementary to at least a portion of an RNA transcript of an MTSP9 polypeptide gene, including a human MTSP9 polypeptide gene. Absolute complementarily is not required.

The amount of MTSP9 polypeptide antisense nucleic acid that is effective in the treatment or prevention of neoplastic disease depends on the nature of the disease, and can be determined empirically by standard clinical techniques. Where possible, it is desirable to determine the antisense cytotoxicity in cells *in vitro*, and then in useful animal model systems prior to testing and use in humans.

**2. RNA interference**

RNA interference (RNAi) (see, *e.g.* Chuang *et al.* (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*:4985) can be employed to inhibit the expression of a gene encoding an MTSP9. Interfering RNA (RNAi) fragments, particularly double-stranded (ds) RNAi, can be used to generate loss-of-MTSP9 function. Methods relating to the use of RNAi to silence genes in organisms including, mammals, C. elegans, Drosophila and plants, and humans are known (see, *e.g.*, Fire *et al.* (1998) *Nature 391*:806-811 Fire (1999) *Trends Genet. 15*:358-363; Sharp (2001) *Genes Dev. 15*:485-490; Hammond, *et al.* (2001) *Nature Rev. Genet.2*:110-1119; Tuschl (2001) *Chem. Biochem. 2*:239-245; Hamilton *et al.* (1999) *Science 286*:950-952; Hammond *et al.* (2000) *Nature 404*:293-296; Zamore *et al.* (2000) *Cell 101*:25-33; Bernstein *et al.* (2001) *Nature 409*: 363-366; Elbashir *et al.* (2001) *Genes Dev. 15*:188-200; Elbashir *et al.* (2001) *Nature 411*:494-498; International PCT application No. WO 01/29058; International PCT application No. WO 99/32619).

Double-stranded RNA (dsRNA)-expressing constructs are introduced into a host, such as an animal or plant using, a replicable vector that remains episomal or integrates into the genome. By selecting appropriate sequences,

nucleic acid (Koller and Smithies, *Proc. Natl. Acad. Sci. USA* 86:8932-8935 (1989); Zijlstra et al., *Nature* 342:435-438 (1989)).

Delivery of the nucleic acid into a patient can be either direct, in which case the patient is directly exposed to the nucleic acid or nucleic acid-carrying vector, or indirect, in which case, cells are first transformed with the nucleic acid *in vitro*, then transplanted into the patient. These two approaches are known, respectively, as *in vivo* or *ex vivo* gene therapy.

In a specific embodiment, the nucleic acid is directly administered *in vivo*, where it is expressed to produce the encoded product. This can be accomplished by any of numerous methods known in the art, *e.g.*, by constructing it as part of an appropriate nucleic acid expression vector and administering it so that it becomes intracellular, *e.g.*, by infection using a defective or attenuated retroviral or other viral vector (see U.S. Patent No. 4,980,286), or by direct injection of naked DNA, or by use of microparticle bombardment (*e.g.*, a gene gun; Biolistic, Dupont), or coating with lipids or cell-surface receptors or transfecting agents, encapsulation in liposomes, microparticles, or microcapsules, or by administering it in linkage to a peptide which is known to enter the nucleus, by administering it in linkage to a ligand subject to receptor-mediated endocytosis (see e.g., Wu and Wu, *J. Biol. Chem.* 262:4429-4432 (1987)) (which can be used to target cell types specifically expressing the receptors), etc. In another embodiment, a nucleic acid-ligand complex can be formed in which the ligand is a fusogenic viral peptide to disrupt endosomes, allowing the nucleic acid to avoid lysosomal degradation. In yet another embodiment, the nucleic acid can be targeted *in vivo* for cell specific uptake and expression, by targeting a specific receptor (see, e.g., PCT Publications WO 92/06180 dated April 16, 1992 (Wu et al.); WO 92/22635 dated December 23, 1992 (Wilson et al.); WO92/20316 dated November 26, 1992 (Findeis et al.); WO93/14188 dated July 22, 1993 (Clarke et al.), WO 93/20221 dated October 14, 1993 (Young)). Alternatively, the nucleic acid can be introduced intracellularly and incorporated within host cell DNA for expression, by homologous recombination (Koller and Smithies, *Proc. Natl. Acad. Sci. USA* 86:8932-8935 (1989); Zijlstra et al., *Nature* 342:435-438 (1989)).

phosphate mediated transfection, or viral infection. Usually, the method of transfer includes the transfer of a selectable marker to the cells. The cells are then placed under selection to isolate those cells that have taken up and are expressing the transferred gene. Those cells are then delivered to a patient.

In this embodiment, the nucleic acid is introduced into a cell prior to administration *in vivo* of the resulting recombinant cell. Such introduction can be carried out by any method known in the art, including but not limited to transfection, electroporation, microinjection, infection with a viral or bacteriophage vector containing the nucleic acid sequences, cell fusion, chromosome-mediated gene transfer, microcell-mediated gene transfer, spheroplast fusion, etc. Numerous techniques are known in the art for the introduction of foreign genes into cells (see *e.g.*, Loeffler and Behr, *Meth. Enzymol.* 217:599-618 (1993); Cohen et al., *Meth. Enzymol.* 217:618-644 (1993); Cline, *Pharmac. Ther.* 29:69-92 (1985)) and can be used, provided that the necessary developmental and physiological functions of the recipient cells are not disrupted. The technique should provide for the stable transfer of the nucleic acid to the cell, so that the nucleic acid is expressible by the cell and generally heritable and expressible by its cell progeny.

The resulting recombinant cells can be delivered to a patient by various methods known in the art. In an embodiment, epithelial cells are injected, *e.g.*, subcutaneously. In another embodiment, recombinant skin cells can be applied as a skin graft onto the patient. Recombinant blood cells (*e.g.*, hematopoietic stem or progenitor cells) can be administered intravenously. The amount of cells envisioned for use depends on the desired effect, patient state, etc., and can be determined by one skilled in the art.

Cells into which a nucleic acid can be introduced for purposes of gene therapy encompass any desired, available cell type, and include but are not limited to epithelial cells, endothelial cells, keratinocytes, fibroblasts, muscle cells, hepatocytes; blood cells such as T lymphocytes, B lymphocytes, monocytes, macrophages, neutrophils, eosinophils, megakaryocytes, granulocytes; various stem or progenitor cells, in particular hematopoietic stem

transplantation immune reactions of the future host/patient. In a particular embodiment, human bone marrow cells can be obtained from the posterior iliac crest by needle aspiration (see, *e.g.*, Kodo et al., *J. Clin. Invest. 73*:1377-1384 (1984)). For example, the HSCs can be made highly enriched or in substantially pure form. This enrichment can be accomplished before, during, or after long-term culturing, and can be done by any techniques known in the art. Long-term cultures of bone marrow cells can be established and maintained by using, for example, modified Dexter cell culture techniques (Dexter et al., *J. Cell Physiol. 91*:335 (1977) or Witlock-Witte culture techniques (Witlock and Witte, *Proc. Natl. Acad. Sci. USA 79*:3608-3612 (1982)).

In a specific embodiment, the nucleic acid to be introduced for purposes of gene therapy includes an inducible promoter operably linked to the coding region, such that expression of the nucleic acid is controllable by controlling the presence or absence of the appropriate inducer of transcription.

**3. Prodrugs**

A method for treating tumors is provided. The method is practiced by administering a prodrug that is cleaved at a specific site by an MTSP9 to release an active drug or precursor that can be converted to active drug *in vivo*. Upon contact with a cell that expresses MTSP9 activity, the prodrug is converted into an active drug. The prodrug can be a conjugate that contains the active agent, such as an anti-tumor drug, such as a cytotoxic agent, or other therapeutic agent (TA), linked to a substrate for the targeted MTSP9, such that the drug or agent is inactive or unable to enter a cell, in the conjugate, but is activated upon cleavage. The prodrug, for example, can contain an oligopeptide, typically a relatively short, less than about 10 amino acids peptide, that is proteolytically cleaved by the targeted MTSP9. Cytotoxic agents, include, but are not limited to, alkylating agents, antiproliferative agents and tubulin binding agents. Others include, vinca drugs, mitomycins, bleomycins and taxanes.

chimeric animal ("knockout animal") in which an MTSP9 polypeptide gene has been inactivated (see Capecchi, *Science* 244:1288-1292 (1989)). The chimeric animal can be bred to produce homozygous knockout animals, which can then be used to produce additional knockout animals. Knockout animals include, but are not limited to, mice, hamsters, sheep, pigs, cattle, and other non-human mammals. For example, a knockout mouse is produced. The resulting animals can serve as models of specific diseases, such as cancers, that exhibit under-expressioof an MTSP9 polypeptide. Such knockout animals can be used as animal models of such diseases *e.g.*, to screen for or test molecules for the ability to treat or prevent such diseases or disorders.

Other types of transgenic animals also can be produced, including those that over-express the MTSP9 polypeptide. Such animals include "knock-in" animals that are animals in which the normal gene is replaced by a variant, such a mutant, an over-expressed form, or other form. For example, one species', such as a rodent's endogenous gene can be replaced by the gene from an other species, such as from a human. Animals also can be produced by non-homologous recombination into other sites in a chromosome; including animals that have a plurality of integration events.

After production of the first generation transgenic animal, a chimeric animal can be bred to produce additional animals with over-expressed or mis-expressed MTSP9 polypeptides. Such animals include, but are not limited to, mice, hamsters, sheep, pigs, cattle and other non-human mammals. The resulting animals can serve as models of specific diseases, such as cancers, that are exhibit over-expression or mis-expression of an MTSP9 polypeptide. Such animals can be used as animal models of such diseases *e.g.*, to screen for or test molecules for the ability to treat or prevent such diseases or disorders. In a specific embodiment, a mouse with over-expressed or mis-expressed MTSP9 polypeptide is produced.

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

composed of RNA extracted from 76 different human tissues (Human Multiple Tissue Expression (MTE) Array; Clontech, Palo Alto, CA; catalog no. 7775-1). The results of this analysis indicate that MTSP9 is highly expressed in the esophagus and expressed at a low level in many other tissues. The MTSP9 transcript is found in kidney (adult and fetal), spleen (adult and fetal), placenta, liver (adult and fetal), thymus, peripheral blood leukocyte, lung (adult and fetal), pancreas, lymph node, bone marrow, trachea, uterus, prostate, esophagus, testes, ovary and the gland organs (mammary, adrenal, thyroid, pituitary and salivary). MTSP9 is also expressed in tumor esophagus tissues, in a lung carcinoma (A549 cell line) and, at a low level, in a colorectal carcinoma (SW480), lymphoma (Raji and Daudi), a cervical carcinoma (HeLaS3) and leukemia (HL-60, K-562 and MOLT-4) cell lines.

**PCR amplification of cDNA encoding full-length protease domain of MTSP9**

To obtain the cDNA fragment encoding the protease domain of MTSP9, an end-to-end PCR amplification using gene-specific primers and the cDNA library from human esophagus was used. The two primers used were: 5′-CGAGTTGTTCCATTAAACGTCAACAGAATAGC-3′ (SEQ ID No. 9) for the 5′ end and 5′-GCATACAGCTTTCTTTGTTTAACTTTTATCGTG -3′ (SEQ ID No. 10) for the 3′ end. The sequences for both primers were derived from the genomic sequence of MTSP9. The 5′ primer contained the sequence that encodes a region immediately upstream of the start of the MTSP9 protease domain (RVVPLNVNRIA; SEQ ID NO. 12). The 3′ primer corresponds to the sequence immediately after the presumed stop codon. A ~750-bp fragment was amplified from the human esophagus cDNA library. The PCR product was isolated and purified using the QIAquick gel extraction kit (Qiagen, Valencia, CA; catalog no. 28704). The MTSP9 PCR product was used to amplify the cDNA fragment containing the appropriate restriction sites for subcloning into the Pichia vector, pPIC9K. The gene-specific primers used were 5′-TCTCTCGAGAAAAGAATAGCATCTGGAGTCATTGCACCCAAG-3′ (SEQ ID No. 13) at the 5′ end and 5′-ATAGCGGCCGCATTAGATGCCTGTTTTTGAAGCAATC-3′ (SEQ ID No. 14) at the 3′ end. The 5′ end primer contained an XhoI site (underlined) immediately

protease, human airway trypsin-like serine protease (Genbank accession number NP004253). Analysis of the MTSP9 protease domain sequence indicated that it is a trypsin-like serine protease domain characterized by the presence of a protease activation cleavage site at the beginning of the domain and the catalytic triad residues (histidine, aspartate and serine) in 3 highly-conserved regions of the catalytic domain. Alignment of the protease domain sequence showed 56% identity with that of endotheliase 1 and 48% identity with the human airway trypsin-like protease domain.

**Sequence analysis**

MTSP9 cDNA and protein sequences were analyzed using MacVector (version 6.5; Oxford Molecular Ltd., Madison, WI). The cDNA encoding the protease domain of MTSP9 is 699 bp long which translates to a 232-amino acid protein. The nucleotide sequence of the protease domain and the translated protein sequence of MTSP9 are as follows (see SEQ ID Nos. 5, 6 and 16):

MTSP9-cDNA containing protease domain sequence Range: 1 to 777 (protease encoding domain 31-729)

```
          10         20         30         40         50         60
AAACGAGTTGTTCCATTAAACGTCAACAGAATAGCATCTGGAGTCATTGCACCCAAGGCG
TTTGCTCAACAAGGTAATTTGCAGTTGTCTTATCGTAGACCTCAGTAACGTGGGTTCCGC
          70         80         90        100        110        120
GCCTGGCCTTGGCAAGCTTCCCTTCAGTATGATAACATCCATCAGTGTGGGGCCACCTTG
CGGACCGGAACCGTTCGAAGGGAAGTCATACTATTGTAGGTAGTCACACCCCGGTGGAAC
         130        140        150        160        170        180
ATTAGTAACACATGGCTTGTCACTGCAGCACACTGCTTCCAGAAGTATAAAAATCCACAT
TAATCATTGTGTACCGAACAGTGACGTCGTGTGACGAAGGTCTTCATATTTTTAGGTGTA
         190        200        210        220        230        240
CAATGGACTGTTAGTTTTGGAACAAAAATCAACCCTCCCTTAATGAAAAGAAATGTCAGA
GTTACCTGACAATCAAAACCTTGTTTTTAGTTGGGAGGGAATTACTTTTCTTTACAGTCT
         250        260        270        280        290        300
AGATTTATTATCCATGAGAAGTACCGCTCTGCAGCAAGAGAGTACGACATTGCTGTTGTG
TCTAAATAATAGGTACTCTTCATGGCGAGACGTCGTTCTCTCATGCTGTAACGACAACAC
         310        320        330        340        350        360
CAGGTCTCTTCCAGAGTCACCTTTTCGGATGACATACGCCGGATTTGTTTGCCAGAAGCC
GTCCAGAGAAGGTCTCAGTGGAAAAGCCTACTGTATGCGGCCTAAACAAACGGTCTTCGG
         370        380        390        400        410        420
TCTGCATCCTTCCAACCAAATTTGACTGTCCACATCACAGGATTTGGAGCACTTTACTAT
AGACGTAGGAAGGTTGGTTTAAACTGACAGGTGTAGTGTCCTAAACCTCGTGAAATGATA
         430        440        450        460        470        480
GGTGGGGAATCCCAAAATGATCTCCGAGAAGCCAGAGTGAAAATCATAAGTGACGATGTC
CCACCCCTTAGGGTTTTACTAGAGGCTCTTCGGTCTCACTTTTAGTATTCACTGCTACAG
         490        500        510        520        530        540
TGCAAGCAACCACAGGTGTATGGCAATGATATAAAACCTGGAATGTTCTGTGCCGGATAT
ACGTTCGTTGGTGTCCACATACCGTTACTATATTTTGGACCTTACAAGACACGGCCTATA
         550        560        570        580        590        600
ATGGAAGGAATTTATGATGCCTGCAGGGGTGATTCTGGGGGACCTTTAGTCACAAGGGAT
TACCTTCCTTAAATACTACGGACGTCCCCACTAAGACCCCCTGGAAATCAGTGTTCCCTA
         610        620        630        640        650        660
CTGAAAGATACGTGGTATCTCATTGGAATTGTAAGCTGGGGAGATAACTGTGGTCAAAAG
```

```
CTGAAAGATACGTGGTATCTCATTGGAATTGTAAGCTGGGGAGATAACTGTGGTCAAAAG
GACTTTCTATGCACCATAGAGTAACCTTAACATTCGACCCCTCTATTGACACCAGTTTTC
 L  K  D  T  W  Y  L  I  G  I  V  S  W  G  D  N  C  G  Q  K

       670        680        690        700        710        720
GACAAGCCTGGAGTCTACACACAAGTGACTTATTACCGAAACTGGATTGCTTCAAAAACA
CTGTTCGGACCTCAGATGTGTGTTCACTGAATAATGGCTTTGACCTAACGAAGTTTTTGT
 D  K  P  G  V  Y  T  Q  V  T  Y  Y  R  N  W  I  A  S  K  T

GGCATCTAA
CCGTAGATT
 G  I  *
```

MTSP9 cDNA and protein sequences were analyzed using MacVector (version 6.5; Oxford Molecular Ltd., Madison, WI). The full-length encoding clone is 1,422 bp long, with a coding region of 1,257 bp long. The translated protein sequence is 418 amino acid residues long. The DNA encoding the protease domain of MTSP9 is 699 bp long which translate to a 232-amino acid protein.

MTSP9 full-length cDNA sequence

Sequence Range: 1 to 1422

```
        10         20         30         40         50         60
GGCGAGCTGGAAGGACGAATCTCCGGCAGCATTCATTACGACAAATGAATGCTGCCGGAT
CCGCTCGACCTTCCTGCTTAGAGGCCGTCGTAAGTAATGCTGTTTACTTACGACGGCCTA

        70         80         90        100        110        120
TAGAAAGTTGAGTTCAGTGGGTGCAGACCTGCAAGATCATATTCTTCCTCCTGTACATGA
ATCTTTCAACTCAAGTCACCCACGTCTGGACGTTCTAGTATAAGAAGGAGGACATGTACT

       130        140        150        160        170        180
TGTATCGGACAGTAGGATTTGGCACCCGAAGCAGAAATCTGAAGCCATGGATGATTGCCG
ACATAGCCTGTCATCCTAAACCGTGGGCTTCGTCTTTAGACTTCGGTACCTACTAACGGC

       190        200        210        220        230        240
TTCTCATTGTGTTGTCCCTGACAGTGGTGGCAGTGACCATAGGTCTCCTGGTTCACTTCC
AAGAGTAACACAACAGGGACTGTCACCACCGTCACTGGTATCCAGAGGACCAAGTGAAGG

       250        260        270        280        290        300
TAGTATTTGACCAAAAAAAGGAGTACTATCATGGCTCCTTTAAAATTTTAGATCCACAAA
ATCATAAACTGGTTTTTTTCCTCATGATAGTACCGAGGAAATTTTAAAATCTAGGTGTTT

       310        320        330        340        350        360
TCAATAACAATTTCGGACAAAGCAACACATATCAACTTAAGGACTTACGAGAGACGACCG
AGTTATTGTTAAAGCCTGTTTCGTTGTGTATAGTTGAATTCCTGAATGCTCTCTGCTGGC

       370        380        390        400        410        420
AAAATTTGGTGGATGAGATATTTATAGATTCAGCCTGGAAGAAAAATTATATCAAGAACC
TTTTAAACCACCTACTCTATAAATATCTAAGTCGGACCTTCTTTTTAATATAGTTCTTGG

       430        440        450        460        470        480
AAGTAGTCAGACTGACTCCAGAGGAAGATGGTGTGAAAGTAGATGTCATTATGGTGTTCC
TTCATCAGTCTGACTGAGGTCTCCTTCTACCACACTTTCATCTACAGTAATACCACAAGG

       490        500        510        520        530        540
AGTTCCCCTCTACTGAACAAAGGGCAGTAAGAGAGAAGAAAATCCAAAGCATCTTAAATC
```

MTSP9 full-length coding sequence
Sequence Range: 1 to 1257

```
          10         20         30         40         50         60
ATGATGTATCGGACAGTAGGATTTGGCACCCGAAGCAGAAATCTGAAGCCATGGATGATT
TACTACATAGCCTGTCATCCTAAACCGTGGGCTTCGTCTTTAGACTTCGGTACCTACTAA

          70         80         90        100        110        120
GCCGTTCTCATTGTGTTGTCCCTGACAGTGGTGGCAGTGACCATAGGTCTCCTGGTTCAC
CGGCAAGAGTAACACAACAGGGACTGTCACCACCGTCACTGGTATCCAGAGGACCAAGTG

         130        140        150        160        170        180
TTCCTAGTATTTGACCAAAAAAAGGAGTACTATCATGGCTCCTTTAAAATTTTAGATCCA
AAGGATCATAAACTGGTTTTTTTCCTCATGATAGTACCGAGGAAATTTTAAAATCTAGGT

         190        200        210        220        230        240
CAAATCAATAACAATTTCGGACAAAGCAACACATATCAACTTAAGGACTTACGAGAGACG
GTTTAGTTATTGTTAAAGCCTGTTTCGTTGTGTATAGTTGAATTCCTGAATGCTCTCTGC

         250        260        270        280        290        300
ACCGAAAATTTGGTGGATGAGATATTTATAGATTCAGCCTGGAAGAAAAATTATATCAAG
TGGCTTTTAAACCACCTACTCTATAAATATCTAAGTCGGACCTTCTTTTTAATATAGTTC

         310        320        330        340        350        360
AACCAAGTAGTCAGACTGACTCCAGAGGAAGATGGTGTGAAAGTAGATGTCATTATGGTG
TTGGTTCATCAGTCTGACTGAGGTCTCCTTCTACCACACTTTCATCTACAGTAATACCAC

         370        380        390        400        410        420
TTCCAGTTCCCCTCTACTGAACAAAGGGCAGTAAGAGAGAAGAAAATCCAAAGCATCTTA
AAGGTCAAGGGGAGATGACTTGTTTCCCGTCATTCTCTCTTCTTTTAGGTTTCGTAGAAT

         430        440        450        460        470        480
AATCAGAAGATAAGGAATTTAAGAGCCTTGCCAATAAATGCCTCATCAGTTCAAGTTAAT
TTAGTCTTCTATTCCTTAAATTCTCGGAACGGTTATTTACGGAGTAGTCAAGTTCAATTA

         490        500        510        520        530        540
GCAATGAGCTCATCAACAGGGGAGTTAACTGTCCAAGCAAGTTGTGGTAAACGAGTTGTT
CGTTACTCGAGTAGTTGTCCCCTCAATTGACAGGTTCGTTCAACACCATTTGCTCAACAA

         550        560        570        580        590        600
CCATTAAACGTCAACAGAATAGCATCTGGAGTCATTGCACCCAAGGCGGCCTGGCCTTGG
GGTAATTTGCAGTTGTCTTATCGTAGACCTCAGTAACGTGGGTTCCGCCGGACCGGAACC

         610        620        630        640        650        660
CAAGCTTCCCTTCAGTATGATAACATCCATCAGTGTGGGGCCACCTTGATTAGTAACACA
GTTCGAAGGGAAGTCATACTATTGTAGGTAGTCACACCCCGGTGGAACTAATCATTGTGT

         670        680        690        700        710        720
TGGCTTGTCACTGCAGCACACTGCTTCCAGAAGTATAAAAATCCACATCAATGGACTGTT
ACCGAACAGTGACGTCGTGTGACGAAGGTCTTCATATTTTTAGGTGTAGTTACCTGACAA

         730        740        750        760        770        780
AGTTTTGGAACAAAAATCAACCCTCCCTTAATGAAAAGAAATGTCAGAAGATTTATTATC
TCAAAACCTTGTTTTTAGTTGGGAGGGAATTACTTTTCTTTACAGTCTTCTAAATAATAG

         790        800        810        820        830        840
CATGAGAAGTACCGCTCTGCAGCAAGAGAGTACGACATTGCTGTTGTGCAGGTCTCTTCC
GTACTCTTCATGGCGAGACGTCGTTCTCTCATGCTGTAACGACAACACGTCCAGAGAAGG

         850        860        870        880        890        900
AGAGTCACCTTTTCGGATGACATACGCCGGATTTGTTTGCCAGAAGCCTCTGCATCCTTC
TCTCAGTGGAAAAGCCTACTGTATGCGGCCTAAACAAACGGTCTTCGGAGACGTAGGAAG
```

## EXAMPLE 2

**Expression of the protease MTSP domains**

Nucleic acid encoding each the MTSP9 and protease domain thereof was cloned into a derivative of the *Pichia pastoris* vector pPIC9K (available from Invitrogen; see SEQ ID NO. 11). Plasmid pPIC9K features include the 5' AOX1 promoter fragment at 1-948; 5' AOX1 primer site at 855-875; alpha-factor secretion signal(s) at 949-1218; alpha-factor primer site at 1152-1172; multiple cloning site at 1192-1241; 3' AOX1 primer site at 1327-1347; 3' AOX1 transcription termination region at 1253-1586; HIS4 ORF at 4514-1980; kanamycin resistance gene at 5743-4928; 3' AOX1 fragment at 6122-6879; ColE1 origin at 7961-7288; and the ampicillin resistance gene at 8966-8106. The plasmid is derived from pPIC9K by eliminating the XhoI site in the kanamycin resistance gene and the resulting vector is herein designated pPIC9Kx.

**C122S mutagenesis of the Protease domain of MTSP9**

The gene encoding the protease domain of MTSP9 was mutagenized by PCR SOE (PCR-based splicing by overlap extension) to replace the unpaired cysteine at position 122 (chymotrypsin numbering system; $Cys_{292}$ in MTSP9) with a serine. Two overlapping gene fragments, each containing the AGT codon for serine at position 122 were PCR amplified using the following primers: for the 5' gene fragment, TCTCTCGAGAAAAGAATAGCATCTGGAGTCATTGCACCC (SEQ ID NO. 13) and AGAGGCTTCTGGCAAACTAATCCGGCGTATGTC (SEQ ID NO. 14); for the 3' gene fragment, ATTCGCGGCCGCTTAGATGCCTGTTTTTGAAGCAAT (SEQ ID NO. 21) and GACATACGCCGGATTAGTTTGCCAGAAGCCTCT (SEQ ID NO. 22). The amplified gene fragments were purified on a 1% agarose gel, mixed and reamplified by PCR to produce the full length coding sequence for the protease domain of MTSP9 C122S. This sequence was then cut with restriction enzymes NotI and XhoI, and ligated into vector pPic9KX.

**MTSP9 fermentation and initial product recovery**

**Fermentation**

sensor was used to control the methanol concentration in the fermenter at that level throughout the induction phase. The methanol fed to the fermenter was supplemented with 2 ml/l PTM4 solution (2.0 g/l $CuSO_4 \bullet 5H_2O$, 0.08 g/l NaI, 3.0 g/l $MnSO_4 \bullet H_2O$, 0.2 g/l $Na_2MoO_4 \bullet 2H_2O$, 0.02 g/l $H_3BO_3$, 0.5 g/l $CoCl_2 \bullet 6H_2O$, 7.0 g/l $ZnCl_2$, 22.0 g/l $FeSO_4 \bullet 7H_2O$, 0.2 g/l biotin, 1.0 ml/l $H_2SO_4$). The induction phase lasted about 42.5 hours.

**Initial Product Recovery**

The supernatant from each of the fermentations was harvested by centrifugation, pooled, and then was concentrated to about 0.5 liter using a 10 kDa ultrafiltration cartridge (A/G Technologies Corp., Needham, MA) on a SRT5 ultrafiltration system (North Carolina SRT, Cary, NC). The concentrate was drained from the system, then the system was rinsed with a volume of 50 mM Hepes, pH 7.0 equal to the concentrated material. The concentrate and the rinse material were combined to yield the final ultrafiltration product of about 1 liter. A final clarification of the supernatant was done with a SartoBran 300 0.45 + 0.2 $\mu$m capsule filter (Sartorius Separations Div., Edgewood, NJ).

**Protein Purification – MTSP9**

Concentrated fermentation supernatant of glycosylated MTSP9 was dialyzed against 50mM HEPES, pH 7.0, filtered and loaded directly onto 147mL SP Sepharose cation exchange column (Amersham-Pharmacia Biotech) which was pre-equilibrated with 50mM HEPES, pH 7.0. Protein was eluted using a linear gradient of 0-500mM NaCl over 7 column volumes at a flow rate of 5mL/min.

Active fractions were pooled then dialyzed against 50mM $Na_2HPO_4$, pH 5.5 overnight. Purified, glycosylated MTSP9 was then deglycosylated by the addition 0.1 ml of Endoglycosidase H (ProZyme, 5 U/ml) per mg of protein and incubating overnight at 4°C with gentle swirling. The dialyzed protein solution was then adjusted to pH 7, filtered and and loaded directly onto 147mL SP Sepharose cation exchange column and eluted as described above. Active fractions were pooled and benzamidine was added to a final concentration of 10 mM. Protein purity was examined by SDS-PAGE and protein concentration

| | |
|---|---|
| S 2266 | H-D-Val-Leu-Arg-pNA.2HCl |
| S 2222 | Bz-Ile-Glu(g-OR)-Gly-Arg-pNA.HCl<br>R = H(50%) and R = $CH_3$(50%) |
| Chromozyme PK | Benzoyl-Pro-Phe-Arg-pNA |
| S 2238 | H-D-Phe-Pip-Arg-pNA.2HCl |
| S 2251 | H-D-Val-Leu-Lys-pNA.2HCl |
| Spectrozyme PI | H-D-Nle-HHT-Lys-pNA.2AcOH |
| | Pyr-Arg-Thr-Lys-Arg-AMC |
| | H-Arg-Gln-Arg-Arg-AMC |
| | Boc-Gln-Gly-Arg-AMC |
| | Z-Arg-Arg-AMC |
| Spectrozyme THE | H-D-HHT-Ala-Arg-pNA.2AcOH |
| Spectrozyme fXIIa | H-D-CHT-Gly-Arg-pNA.2AcOH |
| | CVS 2081-6 ($MeSO_2$-dPhe-Pro-Arg-pNA) |
| | Pefachrome fVIIa ($CH_3SO_2$-D-CHA-But-Arg-pNA) |

pNA = para-nitranilide (chromogenic)
AMC = amino methyl coumarin (fluorescent)

If none of the above substrates are cleaved, a coupled assay, described above, can be used. Briefly, test the ability of the protease to activate and enzyme, such as plasminogen and trypsinogen. To perform these assays, the single chain protease is incubated with a zymogen, such as plasminogen or trypsinogen, in the presence of the a known substrate, such, lys-plasminogen, for the zymogen. If the single chain activates the zymogen, the activated enzyme, such as plasmin and trypsin, will degrade the substrate therefor.

**MTSP-9 Assay for screening modulators**

The protease domain of MTSP9 expressed in *Pichia pastoris* was assayed for inhibition by various test compounds in Costar 96 well tissue culture plates (Corning NY). Approximately 1-20 nM MTSP9 was added without inhibitor, or with 100000 nM inhibitor and 7 1:6 dilutions to 1X direct buffer (29.2 mM Tris, pH 8.4, 29.2 mM Imidazole, 217 mM NaCl (100 $\mu$L final volume)), and allowed to incubate at room temperature for 30 minutes. 400 $\mu$M substrate pefachrome FVIIa (Pentapharm,, Norwalk, CT) was added and the reaction was monitored in a SpectraMAX Plus microplate reader (Molecular Devices, Sunnyvale CA) by following change in absorbance at 405 nm for 20 minutes at 37°C.

initiated by the addition of 50 microliters of rMAP. The final concentrations of all components were identical in both $IC_{50}$ assays (at 30- or 60- and 0-minute).

The initial velocity of chromogenic substrate hydrolysis was measured in both assays by the change of absorbance at 405 nM using a Thermo Max® Kinetic Microplate Reader (Molecular Devices) over a 5 minute period, in which less than 5% of the added substrate was used. The concentration of added inhibitor, which caused a 50% decrease in the initial rate of hydrolysis was defined as the respective $IC_{50}$ value in each of the two assays (30- or 60-minutes and 0-minute).

***In vitro* enzyme assays for specificity determination**

The ability of compounds to act as a selective inhibitor of matriptase activity was assessed by determining the concentration of test compound that inhibits the activity of matriptase by 50%, ($IC_{50}$) as described in the above Example, and comparing $IC_{50}$ value for matriptase to that determined for all or some of the following serine proteases: thrombin, recombinant tissue plasminogen activator (rt-PA), plasmin, activated protein C, chymotrypsin, factor Xa and trypsin.

The buffer used for all assays was HBSA (10 mM HEPES, pH 7.5, 150 mM sodium chloride, 0.1% bovine serum albumin).

The assay for $IC_{50}$ determinations was conducted by combining in appropriate wells of a Corning microtiter plate, 50 microliters of HBSA, 50 microliters of the test compound at a specified concentration (covering a broad concentration range) diluted in HBSA (or HBSA alone for $V_0$ (uninhibited velocity) measurement), and 50 microliters of the enzyme diluted in HBSA. Following a 30 minute incubation at ambient temperature, 50 microliters of the substrate at the concentrations specified below were added to the wells, yielding a final total volume of 200 microliters. The initial velocity of chromogenic substrate hydrolysis was measured by the change in absorbance at 405 nm using a Thermo Max® Kinetic Microplate Reader over a 5 minute period in which less than 5% of the added substrate was used. The concentration of added inhibitor which caused a 50% decrease in the initial rate of hydrolysis was defined as the $IC_{50}$ value.

micromolar (about 3-times Km). Human rt-PA (Activase®) was obtained from Genentech Inc. The enzyme was reconstituted in deionized water and diluted into HBSA prior to the assay in which the final concentration was 1.0 nM.

**Plasmin Assay**

Plasmin catalytic activity was determined using the chromogenic substrate, S-2366 (L-pyroglutamyl-L-prolyl-L-arginine-p-nitroaniline hydrochloride), which was obtained from DiaPharma group. The substrate was made up in deionized water followed by dilution in HBSA prior to the assay in which the final concentration was 300 micromolar (about 2.5-times Km). Purified human plasmin was obtained from Enzyme Research Laboratories, Inc. The enzyme was diluted into HBSA prior to assay in which the final concentration was 1.0 nM.

**Activated Protein C (aPC) Assay**

aPC catalytic activity was determined using the chromogenic substrate, Pefachrome PC (delta-carbobenzloxy-D-lysine-L-prolyl-L-arginine-p-nitroaniline dihydrochloride), obtained from Pentapharm Ltd.). The substrate was made up in deionized water followed by dilution in HBSA prior to the assay in which the final concentration was 400 micromolar (about 3-times Km). Purified human aPC was obtained from Hematologic Technologies, Inc. The enzyme was diluted into HBSA prior to assay in which the final concentration was 1.0 nM.

**Chymotrypsin Assay**

Chymotrypsin catalytic activity was determined using the chromogenic substrate, S-2586 (methoxy-succinyl-L-arginine-L-prolyl-L-tyrosyl-p-nitroanilide), which was obtained from DiaPharma Group. The substrate was made up in deionized water followed by dilution in HBSA prior to the assay in which the final concentration was 100 micromolar (about 9-times Km). Purified (3X-crystallized; CDI) bovine pancreatic alpha-chymotrypsin was obtained from Worthington Biochemical Corp. The enzyme was reconstituted in deionized water and diluted into HBSA prior to assay in which the final concentration was 0.5 nM.

**WHAT IS CLAIMED IS:**

1. A substantially purified single or two chain polypeptide, comprising the protease domain of a type-II membrane-type serine protease 9 (MTSP9) or a catalytically active portion thereof.

2. The polypeptide of claim 1 that is an activated two chain protein.

3. A polypeptide of claim 1 selected from the group consisting of

a polypeptide that comprises a sequence of amino acids encoded by the sequence of nucleotides set forth as nucleotides 31-729 in SEQ ID No. 5 and includes at least about 85% amino acid sequence identity with the seqeuence of amino acids set forth in SEQ ID No. 18;

a polypeptide that comprises a sequence of amino acids encoded by the sequence of nucleotides set forth in SEQ ID No. 17;

a polypeptide that comprises a sequence of amino acids encoded by a sequence of nucleotides that hybridizes along at least 70% of its full-length under conditions of high stringency to the sequence of nucleotides set forth as nucleotides 31-729 in SEQ ID No. 5 or as SEQ ID No. 17;

a polypeptide that comprises the sequence of amino acids set forth as amino acids 11-242 or SEQ ID No. 18;

a polypeptide that comprises a sequence of amino acids having at least about 90% sequence identity with the sequence of amino acids set forth in SEQ ID No. 6 or 18; and

a polypeptide that is encoded by a sequence of nucleotides that is a splice variant of the sequence set forth in SEQ ID No. 17.

4. The polypeptide of claim 1, wherein:

the MTSP9 portion of the polypeptide consists essentially of the protease domain of the MTSP9 or a catalytically active portion thereof.

5. The substantially purified polypeptide of claim 1, wherein the MTSP9 is a human polypeptide.

6. The substantially purified polypeptide of claim 1 that consists essentially of the protease domain of MTSP9 or a catalytically active portion of the protease domain of MTSP9.

16. The polypeptide of claim 15, wherein up to about 10% of the amino acids are replaced with another amino acid.

17. The polypeptide of claim 15, wherein the resulting polypeptide is a single chain or two chain polypeptide and has catalytic activity of at least 50% of the unmutated polypeptide.

18. The polypeptide of claim 15, wherein a free Cysteine in the protease domain is replaced with another amino acid.

19. The polypeptide of claim 18, wherein the replacing amino acid is a serine.

20. An isolated substantially pure polypeptide that consists essentially of the protease domain of MTSP9.

21. A nucleic acid molecule, comprising a sequence of nucleotides that encodes the polypeptide of any of claims 1-20.

22. The nucleic acid molecule of claim 21 that comprises a sequence of nucleotides selected from the group consisting of:

(a) a sequence of nucleotides set forth as nucleotides 31-729 in SEQ ID No. 5 or as SEQ ID No. 17;

(b) a sequence of nucleotides that hybridizes under high stringency along its length or along at least about 70% of the full-length to the sequence of nucleotides set forth as nucleotides 31-729 in SEQ ID No. 5 or as SEQ ID No. 17;

(c) a sequence of nucleotides that encodes the polypeptide of SEQ ID No. 16;

(d) a sequence of nucleotides that is a splice variant of (a, (b) or (c);

(e) a sequence of nucleotides that encodes the protease domain or a catalytically active portion thereof thereof that includes a sequence of nucleotides having at least about 60%, 70%, 80%, 90% or 95% sequence identity the the sequence set forth in SEQ ID Nos. 5, 15 or 17; and and

(f) a sequence of nucleotides comprising degenerate codons of (a), (b), (c), (d) or (e).

23. An isolated nucleic molecule that encodes a mutein of claim 15.

culturing the cell of claim 36 under conditions whereby the encoded polypeptide is expressed by the cell; and

recovering the expressed polypeptide.

43. An antisense nucleic acid molecule that comprises at least 14 contiguous nucleotides or modified nucleotides that are complementary to a contiguous sequence of nucleotides encoding the protease domain of an MTSP9 of claim 1; or

comprises at least 16 contiguous nucleotides or modified nucleotides that are complementary to a contiguous sequence of nucleotides encoding the protease domain of an MTSP9 of any of claims 1-20; or

comprises at least 30 contiguous nucleotides or modified nucleotides that are complementary to a contiguous sequence of nucleotides encoding the protease domain of an MTSP9 of any of claims 1-20, wherein the antisense molecule does include nucleotides 1162-1262 of SEQ ID No. 18.

44. The antisense molecule of claim 43 that includes a contiguous sequence of nucleotides that is the complement of the sequence of nucleoties set forth as nucleotides 31-729 in SEQ ID No. 5 or as SEQ ID No. 17.

45. A double-stranded RNA (dsRNA) molecule that comprises at least about 21 contiguous nucleotides or modified nucleotides from the sequence of nucleotides encoding an MTSP9 of any of claims 1-20.

46. An antibody that specifically binds to the single chain form and/or two-chain form of a protease domain of the polypeptide of any of claims 1-20, or a fragment or derivative of the antibody containing a binding domain thereof, wherein the antibody is a polyclonal antibody or a monoclonal antibody.

47. The antibody of claim 46 that inhibits the enzymatic activity of the polypeptide.

48. An antibody that specifically binds to the single chain form and/or two-chain form of a protease domain of the polypeptide of claim 3, or a fragment or derivative of the antibody containing a binding domain thereof, wherein the antibody is a polyclonal antibody or a monoclonal antibody and it inhibits the enzymatic activity of the polypeptide.

56. A combination, comprising:

an agent or treatment that inhibits the catalytic activity of the polypeptide of any of claims 1-20; and

another treatment or agent selected from anti-tumor and anti-angiogenic treatments and agents.

57. The combination of claim 56, wherein the inhibitor and the anti-tumor and/or anti-angiogenic agent are formulated in a single pharmaceutical composition or each is formulated in separate pharmaceutical compositions.

58. The combination of claim 56, wherein the inhibitor is selected from antibodies and antisense oligonucleotides and double-stranded RNA (dsRNA).

59. A solid support comprising two or more polypeptides of any of claims 1-20 linked thereto either directly or via a linker.

60. The support of claim 59, wherein the polypeptides comprise an array.

61. The support of claim 59, wherein the polypeptides comprise a plurality of different protease domains.

62. A solid support comprising two or more nucleic acid molecules of claim 21 or oligonucleotides portions thereof linked thereto either directly or via a linker, wherein the oligonucleotides contain at least 16 nucleotides.

63. The support of claim 62, wherein the nucleic acid molecules comprise an array.

64. The support of claim 62, wherein the nucleic acid molecules comprise a plurality of molecules that encode different protease domains.

65. A method for identifying compounds that modulate the protease activity of a polypeptide, comprising:

contacting a polypeptide of any of claims 1-20 with a substrate that is proteolytically cleaved by the polypeptide, and, either simultaneously, before or after, adding a test compound or plurality thereof;

measuring the amount of substrate cleaved in the presence of the test compound; and

a polypeptide that comprises a sequence of amino acids encoded by the sequence of nucleotides set forth in SEQ ID No. 17;

a polypeptide that comprises a sequence of amino acids encoded by a sequence of nucleotides that hybridizes under conditions of high stringency to the sequence of nucleotides set forth as nucleotides 31-729 in SEQ ID No. 5 or as SEQ ID No. 17;

a polypeptide that comprises the sequence of amino acids set forth as amino acids 11-242 in SEQ ID No. 16;

a polypeptide that comprises a sequence of amino acids having at least about 60% sequence identity with the sequence of amino acids set forth as amino acids 11-242 in SEQ ID No. 6 or the sequence of amino acids in SEQ ID No. 18; and

a polypeptide that is encoded by a sequence of nucleotides that is a splice variant of the sequence set forth in SEQ ID No. 18.

70. The method of claim 65, wherein the change in the amount of substrate cleaved is assessed by comparing the amount of substrate cleaved in the presence of the test compound with the amount of substrate cleaved in the absence of the test compound.

71. The method of claim 67, wherein a plurality of the polypeptides are linked to a solid support, either directly or via a linker.

72. The method of claim 71, wherein the polypeptides comprise an array.

73. A method of identifying a compound that specifically binds to a single-chain and/or two-chain protease domain and/or to single or two-chain polypeptide and/or to a proteolytically active portion of the single or two chain form thereof of an MTSP9 polypeptide, comprising:

contacting an MTSP9 polypeptide of any of claims 1-20 or a proteolytically active portion thereof with a test compound or plurality thereof under conditions conducive to binding thereof; and either:

a) identifying test compounds that specifically bind to the single chain and/or two chain form of the polypeptide or to a proteolytically active portion of the single an/or two chain form thereof, or

79. A method for identifying activators of the zymogen form of an MTSP9, comprising:

contacting a zymogen form of an MTSP9 polypeptide of any of claims 1-20 or a potentially proteolytically active portion thereof with a substrate of the activated form of the polypeptide;

adding a test compound, wherein the test compound is added before, after or simultaneously with the addition of the substrate; and

detecting cleavage of the substrate, thereby identifying compounds that activate the zymogen.

80. The method of claim 79, wherein the substrate is a chromogenic substrate.

81. The method of claim 79, wherein the substrate is a L-pyroglutamyl-L-prolyl-L-arginine-p-nitroaniline hydrochloride.

82. The method of claim 79, wherein the test compound is a small molecule, a nucleic acid or a polypeptide.

83. A method for treating or preventing a neoplastic disease, in a mammal, comprising administering to a mammal an effective amount of an inhibitor of a polypeptide of any of claims 1-20.

84. The method of claim 83, wherein the inhibitor is an antibody that specifically binds to the polypeptide, or a fragment or derivative of the antibody containing a binding domain thereof, wherein the antibody is a polyclonal antibody or a monoclonal antibody.

85. The method of claim 83, wherein the polypeptide consists essentially of a polypeptide encoded by a sequence of nucleotides that:

(a) a sequence of nucleotides set forth as nucleotides 31-729 in SEQ ID No. 5 or as SEQ ID No. 17;

(b) a sequence of nucleotides that hybridizes under high stringency along its length or along at least about 70% of the full-length to the sequence of nucleotides set forth as nucleotides 31-729 in SEQ ID No. 5 or as SEQ ID No. 17;

(c) a sequence of nucleotides that encodes the polypeptide of SEQ ID No. 16;

90. The method of claim 87, further comprising administering another treatment or agent selected from anti-tumor and anti-angiogenic treatments or agents.

91. The method of claim 87, wherein the polypeptide consists essentially of a polypeptide encoded by:

(a) a sequence of nucleotides set forth as nucleotides 31-729 in SEQ ID No. 5 or as SEQ ID No. 17;

(b) a sequence of nucleotides that hybridizes under high stringency along its length or along at least about 70% of the full-length to the sequence of nucleotides set forth as nucleotides 31-729 in SEQ ID No. 5 or as SEQ ID No. 17;

(c) a sequence of nucleotides that encodes the polypeptide of SEQ ID No. 16;

(d) a sequence of nucleotides that is a splice variant of (a, (b) or (c);

(e) a sequence of nucleotides that encodes the protease domain or a catalytically active portion thereof thereof that includes a sequence of nucleotides having at least about 80% or 85% sequence identity the the sequence set forth in SEQ ID Nos. 5, 15 or 17; and and

(f) a sequence of nucleotides comprising degenerate codons of (a), (b), (c), (d) or (e).

92. The method of claim 87, wherein the polypeptide comprises a polypeptide encoded by a sequence of nucleotides that:

(a) a sequence of nucleotides set forth as nucleotides 31-729 in SEQ ID No. 5 or as SEQ ID No. 17;

(b) a sequence of nucleotides that hybridizes under high stringency along its length to the sequence of nucleotides set forth as nucleotides 31-729 in SEQ ID No. 5 or as SEQ ID No. 17;

(c) a sequence of nucleotides that encodes the polypeptide of SEQ ID No. 16;

(d) a sequence of nucleotides that is a splice variant of (a, (b) or (c); and

97. The method of claim 96, wherein the biological sample is selected from the group consisting of blood, urine, saliva, tears, synovial fluid, sweat, interstitial fluid, cerebrospinal fluid, a sperm sample, ascites fluid, tumor tissue biopsy and circulating tumor cells.

98. The method of claim 96, wherein both forms consist essentially of the protease domain.

99. A method of diagnosing the presence of a pre-malignant lesion, a malignancy, or other pathologic condition in a subject, comprising:

obtaining a biological sample from the subject; and

exposing it to a detectable agent that binds to a two-chain and/or single-chain form of an MTSP9 polypeptide, wherein the pathological condition is characterized by the presence or absence of the two-chain or single-chain form.

100. A method of monitoring tumor progression and/or therapeutic effectiveness, comprising detecting and/or quantifying the level, the form and/or activity of an MTSP9 polypeptide in a body tissue or fluid sample.

101. The method of claim 100, wherein the tumor is a tumor of the breast, cervix, prostate, lung, ovary or colon.

102. The method of claim 100, wherein the body fluid is blood, urine, sweat, saliva, cerebrospinal fluid and synovial fluid.

103. The polypeptide of any of claims 1-20, wherein:

the polypeptide includes at least amino acids 85-87 and/or 160-165 of SEQ ID No. 18.

104. A method for identifying compounds that modulate the protease activity of an MTSP9 polypeptide, comprising:

contacting a polypeptide of any of claims 1-20 or a proteolytically active portion thereof with a substrate that is proteolytically cleaved by the polypeptide, and, either simultaneously, before or after, adding a test compound or plurality thereof;

measuring the amount of substrate cleaved in the presence of the test compound; and

(e) a sequence of nucleotides that encodes the protease domain or a catalytically active portion thereof thereof that includes a sequence of nucleotides having at least about 80% or 85% sequence identity the the sequence set forth in SEQ ID Nos. 5, 15 or 17; and and

(f) a sequence of nucleotides comprising degenerate codons of (a), (b), (c), (d) or (e).

107. The polypeptide of claim 14, wherein the protease domain comprises the sequence of amino acids set forth in SEQ ID No. 16.

108. A transgenic non-human animal, comprising heterologous nucleic acid encoding a polypeptide of any of claims 1-20.

109. A probe or primer that comprises at least 14 contiguous nucleotides or modified nucleotides that are identical to a contiguous sequence of nucleotides encoding the protease domain of an MTSP9 of any of claims 1-20; or

comprises at least 16 contiguous nucleotides or modified nucleotides that are identical to a contiguous sequence of nucleotides encoding the protease domain of an MTSP9 of any of claims 1-20; or

comprises at least 30 contiguous nucleotides or modified nucleotides that are identical to a contiguous sequence of nucleotides encoding the protease domain of an MTSP9 of any of claims 1-20, wherein the antisense molecule does include nucleotides 1162-1262 of SEQ ID No. 18.

SEQUENCE LISTING

<110> Madison, Edwin
Ong, Edgar O.

<120>NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9, THE ENCODED POLYPEPTIDES AND METHODS BASED THEREON

<130> 24745-1615

<140> Not Yet Assigned
<141> Herewith

<150>60/279,228
<151> 03-27-01

<150> 60/291,501
<151> 05-15-01

<160> 22

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 3147
<212> DNA
<213> Homo Sapien

<220>
<221> CDS
<222> (23)...(2589)
<223> Nucleotide sequence encoding MTSP1

<300>
<301> O'Brien, T.J. and Tanimoto, H.
<308> GenBank #AR081724
<309> 2000-08-31
<310> 5,972,616
<311> 1998-02-20
<312> 1999-10-26

<400> 1

```
tcaagagcgg cctcggggta cc atg ggg agc gat cgg gcc cgc aag ggc gga      52
                         Met Gly Ser Asp Arg Ala Arg Lys Gly Gly
                          1               5                  10

ggg ggc ccg aag gac ttc ggc gcg gga ctc aag tac aac tcc cgg cac     100
Gly Gly Pro Lys Asp Phe Gly Ala Gly Leu Lys Tyr Asn Ser Arg His
                 15                  20                  25

gag aaa gtg aat ggc ttg gag gaa ggc gtg gag ttc ctg cca gtc aac     148
Glu Lys Val Asn Gly Leu Glu Glu Gly Val Glu Phe Leu Pro Val Asn
             30                  35                  40

aac gtc aag aag gtg gaa aag cat ggc ccg ggg cgc tgg gtg gtg ctg     196
Asn Val Lys Lys Val Glu Lys His Gly Pro Gly Arg Trp Val Val Leu
         45                  50                  55

gca gcc gtg ctg atc ggc ctc ctc ttg gtc ttg ctg ggg atc ggc ttc     244
```

```
Ala Ala Val Leu Ile Gly Leu Leu Leu Val Leu Leu Gly Ile Gly Phe
     60                  65                  70

ctg gtg tgg cat ttg cag tac cgg gac gtg cgt gtc cag aag gtc ttc      292
Leu Val Trp His Leu Gln Tyr Arg Asp Val Arg Val Gln Lys Val Phe
 75                  80                  85                  90

aat ggc tac atg agg atc aca aat gag aat ttt gtg gat gcc tac gag      340
Asn Gly Tyr Met Arg Ile Thr Asn Glu Asn Phe Val Asp Ala Tyr Glu
                 95                 100                 105

aac tcc aac tcc act gag ttt gta agc ctg gcc agc aag gtg aag gac      388
Asn Ser Asn Ser Thr Glu Phe Val Ser Leu Ala Ser Lys Val Lys Asp
            110                 115                 120

gcg ctg aag ctg ctg tac agc gga gtc cca ttc ctg ggc ccc tac cac      436
Ala Leu Lys Leu Leu Tyr Ser Gly Val Pro Phe Leu Gly Pro Tyr His
        125                 130                 135

aag gag tcg gct gtg acg gcc ttc agc gag ggc agc gtc atc gcc tac      484
Lys Glu Ser Ala Val Thr Ala Phe Ser Glu Gly Ser Val Ile Ala Tyr
    140                 145                 150

tac tgg tct gag ttc agc atc ccg cag cac ctg gtg gag gag gcc gag      532
Tyr Trp Ser Glu Phe Ser Ile Pro Gln His Leu Val Glu Glu Ala Glu
155                 160                 165                 170

cgc gtc atg gcc gag gag cgc gta gtc atg ctg ccc ccg cgg gcg cgc      580
Arg Val Met Ala Glu Glu Arg Val Val Met Leu Pro Pro Arg Ala Arg
                175                 180                 185

tcc ctg aag tcc ttt gtg gtc acc tca gtg gtg gct ttc ccc acg gac      628
Ser Leu Lys Ser Phe Val Val Thr Ser Val Val Ala Phe Pro Thr Asp
            190                 195                 200

tcc aaa aca gta cag agg acc cag gac aac agc tgc agc ttt ggc ctg      676
Ser Lys Thr Val Gln Arg Thr Gln Asp Asn Ser Cys Ser Phe Gly Leu
        205                 210                 215

cac gcc cgc ggt gtg gag ctg atg cgc ttc acc acg ccc ggc ttc cct      724
His Ala Arg Gly Val Glu Leu Met Arg Phe Thr Thr Pro Gly Phe Pro
    220                 225                 230

gac agc ccc tac ccc gct cat gcc cgc tgc cag tgg gcc ctg cgg ggg      772
Asp Ser Pro Tyr Pro Ala His Ala Arg Cys Gln Trp Ala Leu Arg Gly
235                 240                 245                 250

gac gcc gac tca gtg ctg agc ctc acc ttc cgc agc ttt gac ctt gcg      820
Asp Ala Asp Ser Val Leu Ser Leu Thr Phe Arg Ser Phe Asp Leu Ala
                255                 260                 265

tcc tgc gac gag cgc ggc agc gac ctg gtg acg gtg tac aac acc ctg      868
Ser Cys Asp Glu Arg Gly Ser Asp Leu Val Thr Val Tyr Asn Thr Leu
            270                 275                 280

agc ccc atg gag ccc cac gcc ctg gtg cag ttg tgt ggc acc tac cct      916
Ser Pro Met Glu Pro His Ala Leu Val Gln Leu Cys Gly Thr Tyr Pro
        285                 290                 295
```

```
ccc tcc tac aac ctg acc ttc cac tcc tcc cag aac gtc ctg ctc atc      964
Pro Ser Tyr Asn Leu Thr Phe His Ser Ser Gln Asn Val Leu Leu Ile
    300                 305                 310

aca ctg ata acc aac act gag cgg cgg cat ccc ggc ttt gag gcc acc     1012
Thr Leu Ile Thr Asn Thr Glu Arg Arg His Pro Gly Phe Glu Ala Thr
315                 320                 325                 330

ttc ttc cag ctg cct agg atg agc agc tgt gga ggc cgc tta cgt aaa     1060
Phe Phe Gln Leu Pro Arg Met Ser Ser Cys Gly Gly Arg Leu Arg Lys
                335                 340                 345

gcc cag ggg aca ttc aac agc ccc tac tac cca ggc cac tac cca ccc     1108
Ala Gln Gly Thr Phe Asn Ser Pro Tyr Tyr Pro Gly His Tyr Pro Pro
            350                 355                 360

aac att gac tgc aca tgg aac att gag gtg ccc aac aac cag cat gtg     1156
Asn Ile Asp Cys Thr Trp Asn Ile Glu Val Pro Asn Asn Gln His Val
        365                 370                 375

aag gtg agc ttc aaa ttc ttc tac ctg ctg gag ccc ggc gtg cct gcg     1204
Lys Val Ser Phe Lys Phe Phe Tyr Leu Leu Glu Pro Gly Val Pro Ala
    380                 385                 390

ggc acc tgc ccc aag gac tac gtg gag atc aat ggg gag aaa tac tgc     1252
Gly Thr Cys Pro Lys Asp Tyr Val Glu Ile Asn Gly Glu Lys Tyr Cys
395                 400                 405                 410

gga gag agg tcc cag ttc gtc gtc acc agc aac agc aac aag atc aca     1300
Gly Glu Arg Ser Gln Phe Val Val Thr Ser Asn Ser Asn Lys Ile Thr
                415                 420                 425

gtt cgc ttc cac tca gat cag tcc tac acc gac acc ggc ttc tta gct     1348
Val Arg Phe His Ser Asp Gln Ser Tyr Thr Asp Thr Gly Phe Leu Ala
            430                 435                 440

gaa tac ctc tcc tac gac tcc agt gac cca tgc ccg ggg cag ttc acg     1396
Glu Tyr Leu Ser Tyr Asp Ser Ser Asp Pro Cys Pro Gly Gln Phe Thr
        445                 450                 455

tgc cgc acg ggg cgg tgt atc cgg aag gag ctg cgc tgt gat ggc tgg     1444
Cys Arg Thr Gly Arg Cys Ile Arg Lys Glu Leu Arg Cys Asp Gly Trp
    460                 465                 470

gcc gac tgc acc gac cac agc gat gag ctc aac tgc agt tgc gac gcc     1492
Ala Asp Cys Thr Asp His Ser Asp Glu Leu Asn Cys Ser Cys Asp Ala
475                 480                 485                 490

ggc cac cag ttc acg tgc aag aac aag ttc tgc aag ccc ctc ttc tgg     1540
Gly His Gln Phe Thr Cys Lys Asn Lys Phe Cys Lys Pro Leu Phe Trp
                495                 500                 505

gtc tgc gac agt gtg aac gac tgc gga gac aac agc gac gag cag ggg     1588
Val Cys Asp Ser Val Asn Asp Cys Gly Asp Asn Ser Asp Glu Gln Gly
            510                 515                 520

tgc agt tgt ccg gcc cag acc ttc agg tgt tcc aat ggg aag tgc ctc     1636
Cys Ser Cys Pro Ala Gln Thr Phe Arg Cys Ser Asn Gly Lys Cys Leu
        525                 530                 535
```

```
tcg aaa agc cag cag tgc aat ggg aag gac gac tgt ggg gac ggg tcc      1684
Ser Lys Ser Gln Gln Cys Asn Gly Lys Asp Asp Cys Gly Asp Gly Ser
    540                 545                 550

gac gag gcc tcc tgc ccc aag gtg aac gtc gtc act tgt acc aaa cac      1732
Asp Glu Ala Ser Cys Pro Lys Val Asn Val Val Thr Cys Thr Lys His
555                 560                 565                 570

acc tac cgc tgc ctc aat ggg ctc tgc ttg agc aag ggc aac cct gag      1780
Thr Tyr Arg Cys Leu Asn Gly Leu Cys Leu Ser Lys Gly Asn Pro Glu
                575                 580                 585

tgt gac ggg aag gag gac tgt agc gac ggc tca gat gag aag gac tgc      1828
Cys Asp Gly Lys Glu Asp Cys Ser Asp Gly Ser Asp Glu Lys Asp Cys
            590                 595                 600

gac tgt ggg ctg cgg tca ttc acg aga cag gct cgt gtt gtt ggg ggc      1876
Asp Cys Gly Leu Arg Ser Phe Thr Arg Gln Ala Arg Val Val Gly Gly
        605                 610                 615

acg gat gcg gat gag ggc gag tgg ccc tgg cag gta agc ctg cat gct      1924
Thr Asp Ala Asp Glu Gly Glu Trp Pro Trp Gln Val Ser Leu His Ala
    620                 625                 630

ctg ggc cag ggc cac atc tgc ggt gct tcc ctc atc tct ccc aac tgg      1972
Leu Gly Gln Gly His Ile Cys Gly Ala Ser Leu Ile Ser Pro Asn Trp
635                 640                 645                 650

ctg gtc tct gcc gca cac tgc tac atc gat gac aga gga ttc agg tac      2020
Leu Val Ser Ala Ala His Cys Tyr Ile Asp Asp Arg Gly Phe Arg Tyr
                655                 660                 665

tca gac ccc acg cag tgg acg gcc ttc ctg ggc ttg cac gac cag agc      2068
Ser Asp Pro Thr Gln Trp Thr Ala Phe Leu Gly Leu His Asp Gln Ser
            670                 675                 680

cag cgc agc gcc cct ggg gtg cag gag cgc agg ctc aag cgc atc atc      2116
Gln Arg Ser Ala Pro Gly Val Gln Glu Arg Arg Leu Lys Arg Ile Ile
        685                 690                 695

tcc cac ccc ttc ttc aat gac ttc acc ttc gac tat gac atc gcg ctg      2164
Ser His Pro Phe Phe Asn Asp Phe Thr Phe Asp Tyr Asp Ile Ala Leu
    700                 705                 710

ctg gag ctg gag aaa ccg gca gag tac agc tcc atg gtg cgg ccc atc      2212
Leu Glu Leu Glu Lys Pro Ala Glu Tyr Ser Ser Met Val Arg Pro Ile
715                 720                 725                 730

tgc ctg ccg gac gcc tcc cat gtc ttc cct gcc ggc aag gcc atc tgg      2260
Cys Leu Pro Asp Ala Ser His Val Phe Pro Ala Gly Lys Ala Ile Trp
                735                 740                 745

gtc acg ggc tgg gga cac acc cag tat gga ggc act ggc gcg ctg atc      2308
Val Thr Gly Trp Gly His Thr Gln Tyr Gly Gly Thr Gly Ala Leu Ile
            750                 755                 760

ctg caa aag ggt gag atc cgc gtc atc aac cag acc acc tgc gag aac      2356
Leu Gln Lys Gly Glu Ile Arg Val Ile Asn Gln Thr Thr Cys Glu Asn
```

```
        765                 770                 775

ctc ctg ccg cag cag atc acg ccg cgc atg atg tgc gtg ggc ttc ctc     2404
Leu Leu Pro Gln Gln Ile Thr Pro Arg Met Met Cys Val Gly Phe Leu
    780                 785                 790

agc ggc ggc gtg gac tcc tgc cag ggt gat tcc ggg gga ccc ctg tcc     2452
Ser Gly Gly Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Ser
795                 800                 805                 810

agc gtg gag gcg gat ggg cgg atc ttc cag gcc ggt gtg gtg agc tgg     2500
Ser Val Glu Ala Asp Gly Arg Ile Phe Gln Ala Gly Val Val Ser Trp
                815                 820                 825

gga gac ggc tgc gct cag agg aac aag cca ggc gtg tac aca agg ctc     2548
Gly Asp Gly Cys Ala Gln Arg Asn Lys Pro Gly Val Tyr Thr Arg Leu
            830                 835                 840

cct ctg ttt cgg gac tgg atc aaa gag aac act ggg gta ta ggggccgggg   2599
Pro Leu Phe Arg Asp Trp Ile Lys Glu Asn Thr Gly Val
        845                 850                 855

ccacccaaat gtgtacacct gcggggccac ccatcgtcca ccccagtgtg cacgcctgca   2659
ggctggagac tggaccgctg actgcaccag cgcccccaga acatacactg tgaactcaat   2719
ctccagggct ccaaatctgc ctagaaaacc tctcgcttcc tcagcctcca aagtggagct   2779
gggaggtaga aggggaggac actggtggtt ctactgaccc aactgggggc aaaggtttga   2839
agacacagcc tcccccgcca gccccaagct gggccgaggc gcgtttgtgt atatctgcct   2899
cccctgtctg taaggagcag cgggaacgga gcttcggagc ctcctcagtg aaggtggtgg   2959
ggctgccgga tctgggctgt ggggcccttg ggccacgctc ttgaggaagc ccaggctcgg   3019
aggaccctgg aaaacagacg ggtctgagac tgaaattgtt ttaccagctc ccagggtgga   3079
cttcagtgtg tgtatttgtg taaatgggta aaacaattta tttcttttta aaaaaaaaaa   3139
aaaaaaaa                                                           3147

<210> 2
<211> 855
<212> PRT
<213> Homo Sapien

<400> 2
Met Gly Ser Asp Arg Ala Arg Lys Gly Gly Gly Gly Pro Lys Asp Phe
 1               5                  10                  15
Gly Ala Gly Leu Lys Tyr Asn Ser Arg His Glu Lys Val Asn Gly Leu
            20                  25                  30
Glu Glu Gly Val Glu Phe Leu Pro Val Asn Asn Val Lys Lys Val Glu
        35                  40                  45
Lys His Gly Pro Gly Arg Trp Val Val Leu Ala Ala Val Leu Ile Gly
    50                  55                  60
Leu Leu Leu Val Leu Leu Gly Ile Gly Phe Leu Val Trp His Leu Gln
65                  70                  75                  80
Tyr Arg Asp Val Arg Val Gln Lys Val Phe Asn Gly Tyr Met Arg Ile
                85                  90                  95
Thr Asn Glu Asn Phe Val Asp Ala Tyr Glu Asn Ser Asn Ser Thr Glu
            100                 105                 110
Phe Val Ser Leu Ala Ser Lys Val Lys Asp Ala Leu Lys Leu Leu Tyr
        115                 120                 125
Ser Gly Val Pro Phe Leu Gly Pro Tyr His Lys Glu Ser Ala Val Thr
    130                 135                 140
Ala Phe Ser Glu Gly Ser Val Ile Ala Tyr Tyr Trp Ser Glu Phe Ser
145                 150                 155                 160
```

```
Ile Pro Gln His Leu Val Glu Glu Ala Glu Arg Val Met Ala Glu Glu
                165                 170                 175
Arg Val Val Met Leu Pro Pro Arg Ala Arg Ser Leu Lys Ser Phe Val
            180                 185                 190
Val Thr Ser Val Val Ala Phe Pro Thr Asp Ser Lys Thr Val Gln Arg
        195                 200                 205
Thr Gln Asp Asn Ser Cys Ser Phe Gly Leu His Ala Arg Gly Val Glu
    210                 215                 220
Leu Met Arg Phe Thr Thr Pro Gly Phe Pro Asp Ser Pro Tyr Pro Ala
225                 230                 235                 240
His Ala Arg Cys Gln Trp Ala Leu Arg Gly Asp Ala Asp Ser Val Leu
                245                 250                 255
Ser Leu Thr Phe Arg Ser Phe Asp Leu Ala Ser Cys Asp Glu Arg Gly
            260                 265                 270
Ser Asp Leu Val Thr Val Tyr Asn Thr Leu Ser Pro Met Glu Pro His
        275                 280                 285
Ala Leu Val Gln Leu Cys Gly Thr Tyr Pro Pro Ser Tyr Asn Leu Thr
    290                 295                 300
Phe His Ser Ser Gln Asn Val Leu Leu Ile Thr Leu Ile Thr Asn Thr
305                 310                 315                 320
Glu Arg Arg His Pro Gly Phe Glu Ala Thr Phe Phe Gln Leu Pro Arg
                325                 330                 335
Met Ser Ser Cys Gly Gly Arg Leu Arg Lys Ala Gln Gly Thr Phe Asn
            340                 345                 350
Ser Pro Tyr Tyr Pro Gly His Tyr Pro Pro Asn Ile Asp Cys Thr Trp
        355                 360                 365
Asn Ile Glu Val Pro Asn Asn Gln His Val Lys Val Ser Phe Lys Phe
    370                 375                 380
Phe Tyr Leu Leu Glu Pro Gly Val Pro Ala Gly Thr Cys Pro Lys Asp
385                 390                 395                 400
Tyr Val Glu Ile Asn Gly Glu Lys Tyr Cys Gly Glu Arg Ser Gln Phe
                405                 410                 415
Val Val Thr Ser Asn Ser Asn Lys Ile Thr Val Arg Phe His Ser Asp
            420                 425                 430
Gln Ser Tyr Thr Asp Thr Gly Phe Leu Ala Glu Tyr Leu Ser Tyr Asp
        435                 440                 445
Ser Ser Asp Pro Cys Pro Gly Gln Phe Thr Cys Arg Thr Gly Arg Cys
    450                 455                 460
Ile Arg Lys Glu Leu Arg Cys Asp Gly Trp Ala Asp Cys Thr Asp His
465                 470                 475                 480
Ser Asp Glu Leu Asn Cys Ser Cys Asp Ala Gly His Gln Phe Thr Cys
                485                 490                 495
Lys Asn Lys Phe Cys Lys Pro Leu Phe Trp Val Cys Asp Ser Val Asn
            500                 505                 510
Asp Cys Gly Asp Asn Ser Asp Glu Gln Gly Cys Ser Cys Pro Ala Gln
        515                 520                 525
Thr Phe Arg Cys Ser Asn Gly Lys Cys Leu Ser Lys Ser Gln Gln Cys
    530                 535                 540
Asn Gly Lys Asp Asp Cys Gly Asp Gly Ser Asp Glu Ala Ser Cys Pro
545                 550                 555                 560
Lys Val Asn Val Val Thr Cys Thr Lys His Thr Tyr Arg Cys Leu Asn
                565                 570                 575
Gly Leu Cys Leu Ser Lys Gly Asn Pro Glu Cys Asp Gly Lys Glu Asp
            580                 585                 590
Cys Ser Asp Gly Ser Asp Glu Lys Asp Cys Asp Cys Gly Leu Arg Ser
        595                 600                 605
Phe Thr Arg Gln Ala Arg Val Val Gly Gly Thr Asp Ala Asp Glu Gly
    610                 615                 620
Glu Trp Pro Trp Gln Val Ser Leu His Ala Leu Gly Gln Gly His Ile
```

625 630 635 640
Cys Gly Ala Ser Leu Ile Ser Pro Asn Trp Leu Val Ser Ala Ala His
645 650 655
Cys Tyr Ile Asp Asp Arg Gly Phe Arg Tyr Ser Asp Pro Thr Gln Trp
660 665 670
Thr Ala Phe Leu Gly Leu His Asp Gln Ser Gln Arg Ser Ala Pro Gly
675 680 685
Val Gln Glu Arg Arg Leu Lys Arg Ile Ile Ser His Pro Phe Phe Asn
690 695 700
Asp Phe Thr Phe Asp Tyr Asp Ile Ala Leu Leu Glu Leu Glu Lys Pro
705 710 715 720
Ala Glu Tyr Ser Ser Met Val Arg Pro Ile Cys Leu Pro Asp Ala Ser
# 725 730 735
His Val Phe Pro Ala Gly Lys Ala Ile Trp Val Thr Gly Trp Gly His
740 745 750
Thr Gln Tyr Gly Gly Thr Gly Ala Leu Ile Leu Gln Lys Gly Glu Ile
755 760 765
Arg Val Ile Asn Gln Thr Thr Cys Glu Asn Leu Leu Pro Gln Gln Ile
770 775 780
Thr Pro Arg Met Met Cys Val Gly Phe Leu Ser Gly Gly Val Asp Ser
785 790 795 800
Cys Gln Gly Asp Ser Gly Gly Pro Leu Ser Ser Val Glu Ala Asp Gly
805 810 815
Arg Ile Phe Gln Ala Gly Val Val Ser Trp Gly Asp Gly Cys Ala Gln
820 825 830
Arg Asn Lys Pro Gly Val Tyr Thr Arg Leu Pro Leu Phe Arg Asp Trp
835 840 845
Ile Lys Glu Asn Thr Gly Val
850 855

<210> 3
<211> 3147
<212> DNA
<213> Homo Sapien

<220>
<221> CDS
<222> (1865)...(2590)
<223> Nucleic acid sequence of protease domain of MTSP1

<400> 3
tcaagagcgg cctcggggta ccatggggag cgatcgggcc cgcaagggcg gaggggccc 60
gaaggacttc ggcgcgggac tcaagtacaa ctcccggcac gagaaagtga atggcttgga 120
ggaaggcgtg gagttcctgc cagtcaacaa cgtcaagaag gtggaaaagc atggcccggg 180
gcgctgggtg gtgctggcag ccgtgctgat cggcctcctc ttggtcttgc tggggatcgg 240
cttcctggtg tggcatttgc agtaccggga cgtgcgtgtc cagaaggtct tcaatggcta 300
catgaggatc acaaatgaga attttgtgga tgcctacgag aactccaact ccactgagtt 360
tgtaagcctg gccagcaagg tgaaggacgc gctgaagctg ctgtacagcg gagtcccatt 420
cctgggcccc taccacaagg agtcggctgt gacggccttc agcgagggca gcgtcatcgc 480
ctactactgg tctgagttca gcatcccgca gcacctggtg gaggaggccg agcgcgtcat 540
ggccgaggag cgcgtagtca tgctgccccc gcgggcgcgc tccctgaagt cctttgtggt 600
cacctcagtg gtggctttcc ccacggactc caaaacagta cagaggaccc aggacaacag 660
ctgcagcttt ggcctgcacg cccgcggtgt ggagctgatg cgcttcacca cgcccggctt 720
ccctgacagc ccctaccccg ctcatgcccg ctgccagtgg gccctgcggg gggacgccga 780
ctcagtgctg agcctcacct tccgcagctt tgaccttgcg tcctgcgacg agcgcggcag 840
cgacctggtg acggtgtaca acaccctgag ccccatggag ccccacgccc tggtgcagtt 900
gtgtggcacc taccctccct cctacaacct gaccttccac tcctcccaga acgtcctgct 960
catcacactg ataaccaaca ctgagcggcg gcatcccggc tttgaggcca ccttcttcca 1020
gctgcctagg atgagcagct gtggaggccg cttacgtaaa gcccagggga cattcaacag 1080

```
ccctactac ccaggccact acccacccaa cattgactgc acatggaaca ttgaggtgcc   1140
caacaaccag catgtgaagg tgagcttcaa attcttctac ctgctggagc ccggcgtgcc   1200
tgcgggcacc tgccccaagg actacgtgga gatcaatggg gagaaatact gcggagagag   1260
gtcccagttc gtcgtcacca gcaacagcaa caagatcaca gttcgcttcc actcagatca   1320
gtcctacacc gacaccggct tcttagctga atacctctcc tacgactcca gtgacccatg   1380
cccggggcag ttcacgtgcc gcacggggcg gtgtatccgg aaggagctgc gctgtgatgg   1440
ctgggccgac tgcaccgacc acagcgatga gctcaactgc agttgcgacg ccggccacca   1500
gttcacgtgc aagaacaagt tctgcaagcc cctcttctgg gtctgcgaca gtgtgaacga   1560
ctgcggagac aacagcgacg agcaggggtg cagttgtccg gcccagacct tcaggtgttc   1620
caatgggaag tgcctctcga aaagccagca gtgcaatggg aaggacgact gtggggacgg   1680
gtccgacgag gcctcctgcc ccaaggtgaa cgtcgtcact tgtaccaaac acacctaccg   1740
ctgcctcaat gggctctgct tgagcaaggg caaccctgag tgtgacggga aggaggactg   1800
tagcgacggc tcagatgaga aggactgcga ctgtgggctg cggtcattca cgagacaggc   1860
tcgt gtt gtt ggg ggc acg gat gcg gat gag ggc gag tgg ccc tgg cag     1909
     Val Val Gly Gly Thr Asp Ala Asp Glu Gly Glu Trp Pro Trp Gln
      1               5                  10                  15

gta agc ctg cat gct ctg ggc cag ggc cac atc tgc ggt gct tcc ctc     1957
Val Ser Leu His Ala Leu Gly Gln Gly His Ile Cys Gly Ala Ser Leu
                20                  25                  30

atc tct ccc aac tgg ctg gtc tct gcc gca cac tgc tac atc gat gac     2005
Ile Ser Pro Asn Trp Leu Val Ser Ala Ala His Cys Tyr Ile Asp Asp
            35                  40                  45

aga gga ttc agg tac tca gac ccc acg cag tgg acg gcc ttc ctg ggc     2053
Arg Gly Phe Arg Tyr Ser Asp Pro Thr Gln Trp Thr Ala Phe Leu Gly
        50                  55                  60

ttg cac gac cag agc cag cgc agc gcc cct ggg gtg cag gag cgc agg     2101
Leu His Asp Gln Ser Gln Arg Ser Ala Pro Gly Val Gln Glu Arg Arg
    65                  70                  75

ctc aag cgc atc atc tcc cac ccc ttc ttc aat gac ttc acc ttc gac     2149
Leu Lys Arg Ile Ile Ser His Pro Phe Phe Asn Asp Phe Thr Phe Asp
80                  85                  90                  95

tat gac atc gcg ctg ctg gag ctg gag aaa ccg gca gag tac agc tcc     2197
Tyr Asp Ile Ala Leu Leu Glu Leu Glu Lys Pro Ala Glu Tyr Ser Ser
                100                 105                 110

atg gtg cgg ccc atc tgc ctg ccg gac gcc tcc cat gtc ttc cct gcc     2245
Met Val Arg Pro Ile Cys Leu Pro Asp Ala Ser His Val Phe Pro Ala
            115                 120                 125

ggc aag gcc atc tgg gtc acg ggc tgg gga cac acc cag tat gga ggc     2293
Gly Lys Ala Ile Trp Val Thr Gly Trp Gly His Thr Gln Tyr Gly Gly
        130                 135                 140

act ggc gcg ctg atc ctg caa aag ggt gag atc cgc gtc atc aac cag     2341
Thr Gly Ala Leu Ile Leu Gln Lys Gly Glu Ile Arg Val Ile Asn Gln
    145                 150                 155

acc acc tgc gag aac ctc ctg ccg cag cag atc acg ccg cgc atg atg     2389
Thr Thr Cys Glu Asn Leu Leu Pro Gln Gln Ile Thr Pro Arg Met Met
160                 165                 170                 175

tgc gtg ggc ttc ctc agc ggc ggc gtg gac tcc tgc cag ggt gat tcc     2437
Cys Val Gly Phe Leu Ser Gly Gly Val Asp Ser Cys Gln Gly Asp Ser
```

```
                180                 185                 190

ggg gga ccc ctg tcc agc gtg gag gcg gat ggg cgg atc ttc cag gcc      2485
Gly Gly Pro Leu Ser Ser Val Glu Ala Asp Gly Arg Ile Phe Gln Ala
            195                 200                 205

ggt gtg gtg agc tgg gga gac ggc tgc gct cag agg aac aag cca ggc      2533
Gly Val Val Ser Trp Gly Asp Gly Cys Ala Gln Arg Asn Lys Pro Gly
        210                 215                 220

gtg tac aca agg ctc cct ctg ttt cgg gac tgg atc aaa gag aac act      2581
Val Tyr Thr Arg Leu Pro Leu Phe Arg Asp Trp Ile Lys Glu Asn Thr
    225                 230                 235

ggg gta tag gggccggggc cacccaaatg tgtacacctg cggggccacc             2630
Gly Val  *
240

catcgtccac cccagtgtgc acgcctgcag gctggagact ggaccgctga ctgcaccagc   2690
gccccccagaa catacactgt gaactcaatc tccagggctc caaatctgcc tagaaaacct  2750
ctcgcttcct cagcctccaa agtggagctg ggaggtagaa ggggaggaca ctggtggttc   2810

tactgaccca actgggggca aaggtttgaa gacacagcct cccccgccag ccccaagctg   2870
ggccgaggcg cgtttgtgta tatctgcctc ccctgtctgt aaggagcagc gggaacggag   2930
cttcggagcc tcctcagtga aggtggtggg gctgccggat ctgggctgtg gggcccttgg   2990
gccacgctct tgaggaagcc caggctcgga ggaccctgga aaacagacgg gtctgagact   3050
gaaattgttt taccagctcc cagggtggac ttcagtgtgt gtatttgtgt aaatgggtaa   3110
aacaatttat ttctttttaa aaaaaaaaaa aaaaaaa                            3147

<210> 4
<211> 241
<212> PRT
<213> Homo Sapien

<400> 4
Val Val Gly Gly Thr Asp Ala Asp Glu Gly Glu Trp Pro Trp Gln Val
 1               5                  10                  15
Ser Leu His Ala Leu Gly Gln Gly His Ile Cys Gly Ala Ser Leu Ile
            20                  25                  30
Ser Pro Asn Trp Leu Val Ser Ala Ala His Cys Tyr Ile Asp Asp Arg
        35                  40                  45
Gly Phe Arg Tyr Ser Asp Pro Thr Gln Trp Thr Ala Phe Leu Gly Leu
    50                  55                  60
His Asp Gln Ser Gln Arg Ser Ala Pro Gly Val Gln Glu Arg Arg Leu
65                  70                  75                  80
Lys Arg Ile Ile Ser His Pro Phe Phe Asn Asp Phe Thr Phe Asp Tyr
                85                  90                  95
Asp Ile Ala Leu Leu Glu Leu Glu Lys Pro Ala Glu Tyr Ser Ser Met
            100                 105                 110
Val Arg Pro Ile Cys Leu Pro Asp Ala Ser His Val Phe Pro Ala Gly
        115                 120                 125
Lys Ala Ile Trp Val Thr Gly Trp Gly His Thr Gln Tyr Gly Gly Thr
    130                 135                 140
Gly Ala Leu Ile Leu Gln Lys Gly Glu Ile Arg Val Ile Asn Gln Thr
145                 150                 155                 160
Thr Cys Glu Asn Leu Leu Pro Gln Gln Ile Thr Pro Arg Met Met Cys
                165                 170                 175
Val Gly Phe Leu Ser Gly Gly Val Asp Ser Cys Gln Gly Asp Ser Gly
            180                 185                 190
```

```
Gly Pro Leu Ser Ser Val Glu Ala Asp Gly Arg Ile Phe Gln Ala Gly
        195                 200                 205
Val Val Ser Trp Gly Asp Gly Cys Ala Gln Arg Asn Lys Pro Gly Val
    210                 215                 220
Tyr Thr Arg Leu Pro Leu Phe Arg Asp Trp Ile Lys Glu Asn Thr Gly
225                 230                 235                 240
Val
```

<210> 5
<211> 777
<212> DNA
<213> Homo Sapien

<220>
<221> CDS
<222> (1)..(729)
<223> Nucleotide sequence encoding MTSP9, including protease domain (31-729)

<400> 5

```
aaa cga gtt gtt cca tta aac gtc aac aga ata gca tct gga gtc att   48
Lys Arg Val Val Pro Leu Asn Val Asn Arg Ile Ala Ser Gly Val Ile
  1               5                  10                  15

gca ccc aag gcg gcc tgg cct tgg caa gct tcc ctt cag tat gat aac   96
Ala Pro Lys Ala Ala Trp Pro Trp Gln Ala Ser Leu Gln Tyr Asp Asn
            20                  25                  30

atc cat cag tgt ggg gcc acc ttg att agt aac aca tgg ctt gtc act   144
Ile His Gln Cys Gly Ala Thr Leu Ile Ser Asn Thr Trp Leu Val Thr
        35                  40                  45

gca gca cac tgc ttc cag aag tat aaa aat cca cat caa tgg act gtt   192
Ala Ala His Cys Phe Gln Lys Tyr Lys Asn Pro His Gln Trp Thr Val
    50                  55                  60

agt ttt gga aca aaa atc aac cct ccc tta atg aaa aga aat gtc aga   240
Ser Phe Gly Thr Lys Ile Asn Pro Pro Leu Met Lys Arg Asn Val Arg
65                  70                  75                  80

aga ttt att atc cat gag aag tac cgc tct gca gca aga gag tac gac   288
Arg Phe Ile Ile His Glu Lys Tyr Arg Ser Ala Ala Arg Glu Tyr Asp
                85                  90                  95

att gct gtt gtg cag gtc tct tcc aga gtc acc ttt tcg gat gac ata   336
Ile Ala Val Val Gln Val Ser Ser Arg Val Thr Phe Ser Asp Asp Ile
            100                 105                 110

cgc cgg att tgt ttg cca gaa gcc tct gca tcc ttc caa cca aat ttg   384
Arg Arg Ile Cys Leu Pro Glu Ala Ser Ala Ser Phe Gln Pro Asn Leu
        115                 120                 125

act gtc cac atc aca gga ttt gga gca ctt tac tat ggt ggg gaa tcc   432
Thr Val His Ile Thr Gly Phe Gly Ala Leu Tyr Tyr Gly Gly Glu Ser
    130                 135                 140
```

```
caa aat gat ctc cga gaa gcc aga gtg aaa atc ata agt gac gat gtc   480
Gln Asn Asp Leu Arg Glu Ala Arg Val Lys Ile Ile Ser Asp Asp Val
145                 150                 155                 160

tgc aag caa cca cag gtg tat ggc aat gat ata aaa cct gga atg ttc   528
Cys Lys Gln Pro Gln Val Tyr Gly Asn Asp Ile Lys Pro Gly Met Phe
                165                 170                 175

tgt gcc gga tat atg gaa gga att tat gat gcc tgc agg ggt gat tct   576
Cys Ala Gly Tyr Met Glu Gly Ile Tyr Asp Ala Cys Arg Gly Asp Ser
            180                 185                 190

ggg gga cct tta gtc aca agg gat ctg aaa gat acg tgg tat ctc att   624
Gly Gly Pro Leu Val Thr Arg Asp Leu Lys Asp Thr Trp Tyr Leu Ile
        195                 200                 205

gga att gta agc tgg gga gat aac tgt ggt caa aag gac aag cct gga   672
Gly Ile Val Ser Trp Gly Asp Asn Cys Gly Gln Lys Asp Lys Pro Gly
    210                 215                 220

gtc tac aca caa gtg act tat tac cga aac tgg att gct tca aaa aca   720
Val Tyr Thr Gln Val Thr Tyr Tyr Arg Asn Trp Ile Ala Ser Lys Thr
225                 230                 235                 240

ggc atc taa ttcacgataa aagttaaaca aagaaagctg tatgcaggtc atatatgc  777
Gly Ile
```

<210> 6
<211> 242
<212> PRT
<213> Homo sapiens
<223> MTSP9, including protease domain (11-242)

<400> 6

```
Lys Arg Val Val Pro Leu Asn Val Asn Arg Ile Ala Ser Gly Val Ile
  1               5                  10                  15

Ala Pro Lys Ala Ala Trp Pro Trp Gln Ala Ser Leu Gln Tyr Asp Asn
             20                  25                  30

Ile His Gln Cys Gly Ala Thr Leu Ile Ser Asn Thr Trp Leu Val Thr
         35                  40                  45

Ala Ala His Cys Phe Gln Lys Tyr Lys Asn Pro His Gln Trp Thr Val
     50                  55                  60

Ser Phe Gly Thr Lys Ile Asn Pro Pro Leu Met Lys Arg Asn Val Arg
 65                  70                  75                  80

Arg Phe Ile Ile His Glu Lys Tyr Arg Ser Ala Ala Arg Glu Tyr Asp
                 85                  90                  95

Ile Ala Val Val Gln Val Ser Ser Arg Val Thr Phe Ser Asp Asp Ile
            100                 105                 110

Arg Arg Ile Cys Leu Pro Glu Ala Ser Ala Ser Phe Gln Pro Asn Leu
        115                 120                 125
```

Thr Val His Ile Thr Gly Phe Gly Ala Leu Tyr Tyr Gly Gly Glu Ser
130 135 140

Gln Asn Asp Leu Arg Glu Ala Arg Val Lys Ile Ile Ser Asp Asp Val
145 150 155 160

Cys Lys Gln Pro Gln Val Tyr Gly Asn Asp Ile Lys Pro Gly Met Phe
165 170 175

Cys Ala Gly Tyr Met Glu Gly Ile Tyr Asp Ala Cys Arg Gly Asp Ser
180 185 190

Gly Gly Pro Leu Val Thr Arg Asp Leu Lys Asp Thr Trp Tyr Leu Ile
195 200 205

Gly Ile Val Ser Trp Gly Asp Asn Cys Gly Gln Lys Asp Lys Pro Gly
210 215 220

Val Tyr Thr Gln Val Thr Tyr Tyr Arg Asn Trp Ile Ala Ser Lys Thr
225 230 235 240

Gly Ile

<210> 7
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> 5' end primer:

<400> 7
ggcaagcttc ccttcagtat gataacatcc atcag 35

<210> 8
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> 3' end primer

<400> 8
aatgagatac cacgtatctt tcagatccct tg 32

<210> 9
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR 5' end primer

<400> 9
cgagttgttc cattaaacgt caacagaata gc 32

<210> 10
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR 3' end primer

<400> 10
gcatacagct ttctttgttt aacttttatc gtg 33

<210> 11
<211> 9276
<212> DNA
<213> Pichia pastoris

<400> 11
agatctaaca tccaaagacg aaaggttgaa tgaaaccttt ttgccatccg acatccacag 60
gtccattctc acacataagt gccaaacgca acaggagggg atacactagc agcagaccgt 120
tgcaaacgca ggacctccac tcctctttctc ctcaacaccc acttttgcca tcgaaaaacc 180
agcccagtta ttgggcttga ttggagctcg ctcattccaa ttccttctat taggctacta 240
acaccatgac tttattagcc tgtctatcct ggccccccctg gcgaggttca tgtttgttta 300
tttccgaatg caacaagctc cgcattacac ccgaacatca ctccagatga gggctttctg 360
agtgtggggt caaatagttt catgttcccc aaatggccca aaactgacag tttaaacgct 420
gtcttggaac ctaatatgac aaaagcgtga tctcatccaa gatgaactaa gtttggttcg 480
ttgaaatgct aacggccagt tggtcaaaaa gaaacttcca aaagtcgcca taccgtttgt 540
cttgttttggt attgattgac gaatgctcaa aaataatctc attaatgctt agcgcagtct 600
ctctatcgct tctgaacccc ggtgcacctg tgccgaaacg caaatgggga aacacccgct 660
ttttggatga ttatgcattg tctccacatt gtatgcttcc aagattctgg tgggaatact 720
gctgatagcc taacgttcat gatcaaaatt taactgttct aacccctact tgacagcaat 780
atataaacag aaggaagctg ccctgtctta aacctttttt tttatcatca ttattagctt 840
actttcataa ttgcgactgg ttccaattga caagcttttg attttaacga cttttaacga 900
caacttgaga agatcaaaaa acaactaatt attcgaagga tccaaacgat gagatttcct 960
tcaatttta ctgcagtttt attcgcagca tcctccgcat tagctgctcc agtcaacact 1020
acaacagaag atgaaacggc acaaattccg gctgaagctg tcatcggtta ctcagattta 1080
gaaggggatt tcgatgttgc tgttttgcca ttttccaaca gcacaaataa cgggttattg 1140
tttataaata ctactattgc cagcattgct gctaaagaag aaggggtatc tctcgagaaa 1200
agagaggctg aagcttacgt agaattccct agggcggccg cgaattaatt cgccttagac 1260
atgactgttc ctcagttcaa gttgggcact tacgagaaga ccggtcttgc tagattctaa 1320
tcaagaggat gtcagaatgc catttgcctg agagatgcag gcttcatttt tgatactttt 1380
ttatttgtaa cctatatagt ataggatttt ttttgtcatt ttgtttcttc tcgtacgagc 1440
ttgctcctga tcagcctatc tcgcagctga tgaatatctt gtggtagggg tttgggaaaa 1500
tcattcgagt ttgatgtttt tcttggtatt tcccactcct cttcagagta cagaagatta 1560
agtgagaagt tcgtttgtgc aagcttatcg ataagcttta atgcggtagt ttatcacagt 1620
taaattgcta acgcagtcag gcaccgtgta tgaaatctaa caatgcgctc atcgtcatcc 1680
tcggcaccgt caccctggat gctgtaggca taggcttggt tatgccggta ctgccgggcc 1740
tcttgcggga tatcgtccat tccgacagca tcgccagtca ctatggcgtg ctgctagcgc 1800
tatatgcgtt gatgcaattt ctatgcgcac ccgttctcgg agcactgtcc gaccgctttg 1860
gccgccgccc agtcctgctc gcttcgctac ttggagccac tatcgactac gcgatcatgg 1920
cgaccacacc cgtcctgtgg atctatcgaa tctaaatgta agttaaaatc tctaaataat 1980
taaataagtc ccagtttctc catacgaacc ttaacagcat tgcggtgagc atctagacct 2040
tcaacagcag ccagatccat cactgcttgg ccaatatgtt tcagtccctc aggagttacg 2100
tcttgtgaag tgatgaactt ctggaaggtt gcagtgttaa ctccgctgta ttgacgggca 2160
tatccgtacg ttggcaaagt gtggttggta ccggaggagt aatctccaca actctctgga 2220
gagtaggcac caacaaacac agatccagcg tgttgtactt gatcaacata agaagaagca 2280
ttctcgattt gcaggatcaa gtgttcagga gcgtactgat tggacatttc caaagcctgc 2340
tcgtaggttg caaccgatag ggttgtagag tgtgcaatac acttgcgtac aatttcaacc 2400
cttggcaact gcacagcttg gttgtgaaca gcatcttcaa ttctggcaag ctccttgtct 2460
gtcatatcga cagccaacag aatcacctgg gaatcaatac catgttcagc ttgagacaga 2520

aggtctgagg caacgaaatc tggatcagcg tatttatcag caataactag aacttcagaa 2580
ggcccagcag gcatgtcaat actacacagg gctgatgtgt cattttgaac catcatcttg 2640
gcagcagtaa cgaactggtt tcctggacca aatattttgt cacacttagg aacagtttct 2700
gttccgtaag ccatagcagc tactgcctgg gcgcctcctg ctagcacgat acacttagca 2760
ccaaccttgt gggcaacgta gatgacttct ggggtaaggg taccatcctt cttaggtgga 2820
gatgcaaaaa caatttcttt gcaaccagca actttggcag gaacacccag catcagggaa 2880
gtggaaggca gaattgcggt tccaccagga atatagaggc caactttctc aataggtctt 2940
gcaaaacgag agcagactac accagggcaa gtctcaactt gcaacgtctc cgttagttga 3000
gcttcatgga atttcctgac gttatctata gagagatcaa tggctctctt aacgttatct 3060
ggcaattgca taagttcctc tgggaaagga gcttctaaca caggtgtctt caaagcgact 3120
ccatcaaact tggcagttag ttctaaaagg gctttgtcac cattttgacg aacattgtcg 3180
acaattggtt tgactaattc cataatctgt tccgttttct ggataggacg acgaagggca 3240
tcttcaattt cttgtgagga ggccttagaa acgtcaattt tgcacaattc aatacgacct 3300
tcagaaggga cttctttagg tttggattct tctttaggtt gttccttggt gtatcctggc 3360
ttggcatctc ctttccttct agtgaccttt agggacttca tatccaggtt tctctccacc 3420
tcgtccaacg tcacaccgta cttggcacat ctaactaatg caaaataaaa taagtcagca 3480
cattcccagg ctatatcttc cttggattta gcttctgcaa gttcatcagc ttcctcccta 3540
attttagcgt tcaacaaaac ttcgtcgtca aataaccgtt tggtataaga accttctgga 3600
gcattgctct tacgatccca caaggtggct tccatggctc taagaccctt tgattggcca 3660
aaacaggaag tgcgttccaa gtgacagaaa ccaacacctg tttgttcaac cacaaatttc 3720
aagcagtctc catcacaatc caattcgata cccagcaact tttgagttgc tccagatgta 3780
gcacctttat accacaaacc gtgacgacga gattggtaga ctccagtttg tgtccttata 3840
gcctccggaa tagacttttt ggacgagtac accaggccca acgagtaatt agaagagtca 3900
gccaccaaag tagtgaatag accatcgggg cggtcagtag tcaaagacgc caacaaaatt 3960
tcactgacag ggaacttttt gacatcttca gaaagttcgt attcagtagt caattgccga 4020
gcatcaataa tggggattat accagaagca acagtggaag tcacatctac caactttgcg 4080
gtctcagaaa aagcataaac agttctacta ccgccattag tgaaactttt caaatcgccc 4140
agtggagaag aaaaaggcac agcgatacta gcattagcgg gcaaggatgc aactttatca 4200
accagggtcc tatagataac cctagcgcct gggatcatcc tttggacaac tctttctgcc 4260
aaatctaggt ccaaaatcac ttcattgata ccattattgt acaacttgag caagttgtcg 4320
atcagctcct caaattggtc ctctgtaacg gatgactcaa cttgcacatt aacttgaagc 4380
tcagtcgatt gagtgaactt gatcaggttg tgcagctggt cagcagcata gggaaacacg 4440
gcttttccta ccaaactcaa ggaattatca aactctgcaa cacttgcgta tgcaggtagc 4500
aagggaaatg tcatacttga agtcggacag tgagtgtagt cttgagaaat tctgaagccg 4560
tatttttatt atcagtgagt cagtcatcag gagatcctct acgccggacg catcgtggcc 4620
gacctgcagg gggggggggg gcgctgaggt ctgcctcgtg aagaaggtgt tgctgactca 4680
taccaggcct gaatcgcccc atcatccagc cagaaagtga gggagccacg gttgatgaga 4740
gctttgttgt aggtggacca gttggtgatt ttgaactttt gctttgccac ggaacggtct 4800
gcgttgtcgg gaagatgcgt gatctgatcc ttcaactcag caaaagttcg atttattcaa 4860
caaagccgcc gtcccgtcaa gtcagcgtaa tgctctgcca gtgttacaac caattaacca 4920
attctgatta gaaaaactca tcgagcatca aatgaaactg caatttattc atatcaggat 4980
tatcaatacc atatttttga aaaagccgtt tctgtaatga aggagaaaac tcaccgaggc 5040
agttccatag gatggcaaga tcctggtatc ggtctgcgat tccgactcgt ccaacatcaa 5100
tacaacctat taatttcccc tcgtcaaaaa taaggttatc aagtgagaaa tcaccatgag 5160
tgacgactga atccggtgag aatggcaaaa gcttatgcat ttctttccag acttgttcaa 5220
caggccagcc attacgctcg tcatcaaaat cactcgcatc aaccaaaccg ttattcattc 5280
gtgattgcgc ctgagcgaga cgaaatacgc gatcgctgtt aaaaggacaa ttacaaacag 5340
gaatcgaatg caaccggcgc aggaacactg ccagcgcatc aacaatattt tcacctgaat 5400
caggatattc ttctaatacc tggaatgctg ttttcccggg gatcgcagtg gtgagtaacc 5460
atgcatcatc aggagtacgg ataaaatgct tgatggtcgg aagaggcata aattccgtca 5520
gccagtttag tctgaccatc tcatctgtaa catcattggc aacgctacct ttgccatgtt 5580
tcagaaacaa ctctggcgca tcgggcttcc catacaatcg atagattgtc gcacctgatt 5640
gcccgacatt atcgcgagcc catttatacc catataaatc agcatccatg ttggaattta 5700
atcgcggcct cgagcaagac gtttcccgtt gaatatggct cataacaccc cttgtattac 5760
tgtttatgta agcagacagt tttattgttc atgatgatat atttttatct tgtgcaatgt 5820
aacatcagag attttgagac acaacgtggc tttcccccccc ccccctgcag gtcggcatca 5880
ccggcgccac aggtgcggtt gctggcgcct atatcgccga catcaccgat ggggaagatc 5940
gggctcgcca cttcgggctc atgagcgctt gtttcggcgt gggtatggtg gcaggccccg 6000
tggccggggg actgttgggc gccatctcct tgcatgcacc attccttgcg gcggcggtgc 6060

tcaacggcct caacctacta ctgggctgct tcctaatgca ggagtcgcat aagggagagc 6120
gtcgagtatc tatgattgga agtatgggaa tggtgatacc cgcattcttc agtgtcttga 6180
ggtctcctat cagattatgc ccaactaaag caaccggagg aggagatttc atggtaaatt 6240
tctctgactt ttggtcatca gtagactcga actgtgagac tatctcggtt atgacagcag 6300
aaatgtcctt cttggagaca gtaaatgaag tcccaccaat aaagaaatcc ttgttatcag 6360
gaacaaactt cttgtttcga acttttttcgg tgccttgaac tataaaatgt agagtggata 6420
tgtcgggtag gaatggagcg ggcaaatgct taccttctgg accttcaaga ggtatgtagg 6480
gtttgtagat actgatgcca acttcagtga caacgttgct atttcgttca aaccattccg 6540
aatccagaga aatcaaagtt gtttgtctac tattgatcca agccagtgcg gtcttgaaac 6600
tgacaatagt gtgctcgtgt tttgaggtca tctttgtatg aataaatcta gtctttgatc 6660
taaataatct tgacgagcca aggcgataaa tacccaaatc taaaactctt ttaaaacgtt 6720
aaaaggacaa gtatgtctgc ctgtattaaa ccccaaatca gctcgtagtc tgatcctcat 6780
caacttgagg ggcactatct tgttttagag aaatttgcgg agatgcgata tcgagaaaaa 6840
ggtacgctga ttttaaacgt gaaatttatc tcaagatctc tgcctcgcgc gtttcggtga 6900
tgacggtgaa aacctctgac acatgcagct cccggagacg gtcacagctt gtctgtaagc 6960
ggatgccggg agcagacaag cccgtcaggg cgcgtcagcg ggtgttggcg ggtgtcgggg 7020
cgcagccatg acccagtcac gtagcgatag cggagtgtat actggcttaa ctatgcggca 7080
tcagagcaga ttgtactgag agtgcaccat atgcggtgtg aaataccgca cagatgcgta 7140
aggagaaaat accgcatcag gcgctcttcc gcttcctcgc tcactgactc gctgcgctcg 7200
gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatacg gttatccaca 7260
gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac 7320
cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gccccctga cgagcatcac 7380
aaaaatcgac gctcaagtca gaggtggcga aacccgacag gactataaag ataccaggcg 7440
tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgccgct taccggatac 7500
ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc aatgctcacg ctgtaggtat 7560
ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag 7620
cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt aagacacgac 7680
ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcggt 7740
gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttggt 7800
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 7860
aaacaaacca ccgctggtag cggtggtttt tttgtttgca agcagcagat tacgcgcaga 7920
aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggaac 7980
gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt cacctagatc 8040
cttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct 8100
gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca 8160
tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct 8220
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca 8280
ataaaccagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc 8340
atccagtcta ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg 8400
cgcaacgttg ttgccattgc tgcaggcatc gtggtgtcac gctcgtcgtt tggtatggct 8460
tcattcagct ccggttccca acgatcaagg cgagttacat gatccccat gttgtgcaaa 8520
aaagcggtta gctccttcgg tcctccgatc gttgtcagaa gtaagttggc cgcagtgtta 8580
tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc 8640
ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg 8700
agttgctctt gcccggcgtc aacacgggat aataccgcgc cacatagcag aactttaaaa 8760
gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg 8820
agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc 8880
accagcgttt ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg 8940
gcgacacgga aatgttgaat actcatactc ttccttttc aatattattg aagcatttat 9000
cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata 9060
ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tctaagaaac cattattatc 9120
atgacattaa cctataaaaa taggcgtatc acgaggccct ttcgtcttca agaattaatt 9180
ctcatgtttg acagcttatc atcgataagc tgactcatgt tggtattgtg aaatagacgc 9240
agatcgggaa cactgaaaaa taacagttat tattcg 9276

<210> 12
<211> 11
<212> PRT

<213> MTSP9 region adjacent to and including the cleavage site starting the protease domain (Arg↓Ile)

<400> 12
Arg Val Val Pro Leu Asn Val Asn Arg Ile Ala
1 5 10

<210> 13
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 13
tctctcgaga aaagaatagc atctggagtc attgcaccc 39

<210> 14
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 14
agaggcttct ggcaaactaa tcggcgtatg tc 32

<210> 15
<211> 11
<212> PRT
<213> Pichia protease cleavage site
<400> 15

Lys Arg Ile Ala Ser Gly Val Ile Ala Pro Lys
1 5 10

<210> 16

<211> 232
<212> PRT
<213> Homo sapiens
<223> Protease domain of MTSP9 (1-232)

<400> 16

Ile Ala Ser Gly Val Ile Ala Pro Lys Ala Ala Trp Pro Trp Gln Ala
1 5 10 15

Ser Leu Gln Tyr Asp Asn Ile His Gln Cys Gly Ala Thr Leu Ile Ser
20 25 30

Asn Thr Trp Leu Val Thr Ala Ala His Cys Phe Gln Lys Tyr Lys Asn
35 40 45

Pro His Gln Trp Thr Val Ser Phe Gly Thr Lys Ile Asn Pro Pro Leu

```
     50                  55                  60

Met Lys Arg Asn Val Arg Arg Phe Ile Ile His Glu Lys Tyr Arg Ser
 65                  70                  75                  80

Ala Ala Arg Glu Tyr Asp Ile Ala Val Val Gln Val Ser Ser Arg Val
                85                  90                  95

Thr Phe Ser Asp Asp Ile Arg Arg Ile Cys Leu Pro Glu Ala Ser Ala
            100                 105                 110

Ser Phe Gln Pro Asn Leu Thr Val His Ile Thr Gly Phe Gly Ala Leu
        115                 120                 125

Tyr Tyr Gly Gly Glu Ser Gln Asn Asp Leu Arg Glu Ala Arg Val Lys
    130                 135                 140

Ile Ile Ser Asp Asp Val Cys Lys Gln Pro Gln Val Tyr Gly Asn Asp
145                 150                 155                 160

Ile Lys Pro Gly Met Phe Cys Ala Gly Tyr Met Glu Gly Ile Tyr Asp
                165                 170                 175

Ala Cys Arg Gly Asp Ser Gly Gly Pro Leu Val Thr Arg Asp Leu Lys
            180                 185                 190

Asp Thr Trp Tyr Leu Ile Gly Ile Val Ser Trp Gly Asp Asn Cys Gly
        195                 200                 205

Gln Lys Asp Lys Pro Gly Val Tyr Thr Gln Val Thr Tyr Tyr Arg Asn
    210                 215                 220

Trp Ile Ala Ser Lys Thr Gly Ile
225                 230
```

<210> 17
<211> 1257
<212> DNA
<213> Homo sapien

<220>
<221> CDS
<222> (1)...(1257)
<223> Nucleotide sequence encoding full-length MTSP9

<400> 17

```
atg atg tat cgg aca gta gga ttt ggc acc cga agc aga aat ctg aag      48
Met Met Tyr Arg Thr Val Gly Phe Gly Thr Arg Ser Arg Asn Leu Lys
 1               5                  10                  15

cca tgg atg att gcc gtt ctc att gtg ttg tcc ctg aca gtg gtg gca      96
Pro Trp Met Ile Ala Val Leu Ile Val Leu Ser Leu Thr Val Val Ala
            20                  25                  30

gtg acc ata ggt ctc ctg gtt cac ttc cta gta ttt gac caa aaa aag     144
Val Thr Ile Gly Leu Leu Val His Phe Leu Val Phe Asp Gln Lys Lys
        35                  40                  45
```

```
gag tac tat cat ggc tcc ttt aaa att tta gat cca caa atc aat aac      192
Glu Tyr Tyr His Gly Ser Phe Lys Ile Leu Asp Pro Gln Ile Asn Asn
     50                  55                  60

aat ttc gga caa agc aac aca tat caa ctt aag gac tta cga gag acg      240
Asn Phe Gly Gln Ser Asn Thr Tyr Gln Leu Lys Asp Leu Arg Glu Thr
 65                  70                  75                  80

acc gaa aat ttg gtg gat gag ata ttt ata gat tca gcc tgg aag aaa      288
Thr Glu Asn Leu Val Asp Glu Ile Phe Ile Asp Ser Ala Trp Lys Lys
                 85                  90                  95

aat tat atc aag aac caa gta gtc aga ctg act cca gag gaa gat ggt      336
Asn Tyr Ile Lys Asn Gln Val Val Arg Leu Thr Pro Glu Glu Asp Gly
            100                 105                 110

gtg aaa gta gat gtc att atg gtg ttc cag ttc ccc tct act gaa caa      384
Val Lys Val Asp Val Ile Met Val Phe Gln Phe Pro Ser Thr Glu Gln
        115                 120                 125

agg gca gta aga gag aag aaa atc caa agc atc tta aat cag aag ata      432
Arg Ala Val Arg Glu Lys Lys Ile Gln Ser Ile Leu Asn Gln Lys Ile
    130                 135                 140

agg aat tta aga gcc ttg cca ata aat gcc tca tca gtt caa gtt aat      480
Arg Asn Leu Arg Ala Leu Pro Ile Asn Ala Ser Ser Val Gln Val Asn
145                 150                 155                 160

gca atg agc tca tca aca ggg gag tta act gtc caa gca agt tgt ggt      528
Ala Met Ser Ser Ser Thr Gly Glu Leu Thr Val Gln Ala Ser Cys Gly
                165                 170                 175

aaa cga gtt gtt cca tta aac gtc aac aga ata gca tct gga gtc att      576
Lys Arg Val Val Pro Leu Asn Val Asn Arg Ile Ala Ser Gly Val Ile
            180                 185                 190

gca ccc aag gcg gcc tgg cct tgg caa gct tcc ctt cag tat gat aac      624
Ala Pro Lys Ala Ala Trp Pro Trp Gln Ala Ser Leu Gln Tyr Asp Asn
        195                 200                 205

atc cat cag tgt ggg gcc acc ttg att agt aac aca tgg ctt gtc act      672
Ile His Gln Cys Gly Ala Thr Leu Ile Ser Asn Thr Trp Leu Val Thr
    210                 215                 220

gca gca cac tgc ttc cag aag tat aaa aat cca cat caa tgg act gtt      720
Ala Ala His Cys Phe Gln Lys Tyr Lys Asn Pro His Gln Trp Thr Val
225                 230                 235                 240

agt ttt gga aca aaa atc aac cct ccc tta atg aaa aga aat gtc aga      768
Ser Phe Gly Thr Lys Ile Asn Pro Pro Leu Met Lys Arg Asn Val Arg
                245                 250                 255

aga ttt att atc cat gag aag tac cgc tct gca gca aga gag tac gac      816
Arg Phe Ile Ile His Glu Lys Tyr Arg Ser Ala Ala Arg Glu Tyr Asp
            260                 265                 270

att gct gtt gtg cag gtc tct tcc aga gtc acc ttt tcg gat gac ata      864
Ile Ala Val Val Gln Val Ser Ser Arg Val Thr Phe Ser Asp Asp Ile
        275                 280                 285
```

```
cgc cgg att tgt ttg cca gaa gcc tct gca tcc ttc caa cca aat ttg      912
Arg Arg Ile Cys Leu Pro Glu Ala Ser Ala Ser Phe Gln Pro Asn Leu
    290                 295                 300

act gtc cac atc aca gga ttt gga gca ctt tac tat ggt ggg gaa tcc      960
Thr Val His Ile Thr Gly Phe Gly Ala Leu Tyr Tyr Gly Gly Glu Ser
305                 310                 315                 320

caa aat gat ctc cga gaa gcc aga gtg aaa atc ata agt gac gat gtc     1008
Gln Asn Asp Leu Arg Glu Ala Arg Val Lys Ile Ile Ser Asp Asp Val
                325                 330                 335

tgc aag caa cca cag gtg tat ggc aat gat ata aaa cct gga atg ttc     1056
Cys Lys Gln Pro Gln Val Tyr Gly Asn Asp Ile Lys Pro Gly Met Phe
            340                 345                 350

tgt gcc gga tat atg gaa gga att tat gat gcc tgc agg ggt gat tct     1104
Cys Ala Gly Tyr Met Glu Gly Ile Tyr Asp Ala Cys Arg Gly Asp Ser
        355                 360                 365

ggg gga cct tta gtc aca agg gat ctg aaa gat acg tgg tat ctc att     1152
Gly Gly Pro Leu Val Thr Arg Asp Leu Lys Asp Thr Trp Tyr Leu Ile
    370                 375                 380

gga att gta agc tgg gga gat aac tgt ggt caa aag gac aag cct gga     1200
Gly Ile Val Ser Trp Gly Asp Asn Cys Gly Gln Lys Asp Lys Pro Gly
385                 390                 395                 400

gtc tac aca caa gtg act tat tac cga aac tgg att gct tca aaa aca     1248
Val Tyr Thr Gln Val Thr Tyr Tyr Arg Asn Trp Ile Ala Ser Lys Thr
                405                 410                 415

ggc atc taa                                                        1257
Gly Ile  *
```

<210> 18
<211> 418
<212> PRT
<213> Homo sapien

<400> 18

```
Met Met Tyr Arg Thr Val Gly Phe Gly Thr Arg Ser Arg Asn Leu Lys
 1               5                  10                  15
Pro Trp Met Ile Ala Val Leu Ile Val Leu Ser Leu Thr Val Val Ala
            20                  25                  30
Val Thr Ile Gly Leu Leu Val His Phe Leu Val Phe Asp Gln Lys Lys
        35                  40                  45
Glu Tyr Tyr His Gly Ser Phe Lys Ile Leu Asp Pro Gln Ile Asn Asn
    50                  55                  60
Asn Phe Gly Gln Ser Asn Thr Tyr Gln Leu Lys Asp Leu Arg Glu Thr
65                  70                  75                  80
Thr Glu Asn Leu Val Asp Glu Ile Phe Ile Asp Ser Ala Trp Lys Lys
                85                  90                  95
Asn Tyr Ile Lys Asn Gln Val Val Arg Leu Thr Pro Glu Glu Asp Gly
            100                 105                 110
Val Lys Val Asp Val Ile Met Val Phe Gln Phe Pro Ser Thr Glu Gln
```

```
        115                 120                 125
Arg Ala Val Arg Glu Lys Lys Ile Gln Ser Ile Leu Asn Gln Lys Ile
    130                 135                 140
Arg Asn Leu Arg Ala Leu Pro Ile Asn Ala Ser Ser Val Gln Val Asn
145                 150                 155                 160
Ala Met Ser Ser Ser Thr Gly Glu Leu Thr Val Gln Ala Ser Cys Gly
                165                 170                 175
Lys Arg Val Val Pro Leu Asn Val Asn Arg Ile Ala Ser Gly Val Ile
            180                 185                 190
Ala Pro Lys Ala Ala Trp Pro Trp Gln Ala Ser Leu Gln Tyr Asp Asn
        195                 200                 205
Ile His Gln Cys Gly Ala Thr Leu Ile Ser Asn Thr Trp Leu Val Thr
    210                 215                 220
Ala Ala His Cys Phe Gln Lys Tyr Lys Asn Pro His Gln Trp Thr Val
225                 230                 235                 240
Ser Phe Gly Thr Lys Ile Asn Pro Pro Leu Met Lys Arg Asn Val Arg
                245                 250                 255
Arg Phe Ile Ile His Glu Lys Tyr Arg Ser Ala Ala Arg Glu Tyr Asp
            260                 265                 270
Ile Ala Val Val Gln Val Ser Ser Arg Val Thr Phe Ser Asp Asp Ile
        275                 280                 285
Arg Arg Ile Cys Leu Pro Glu Ala Ser Ala Ser Phe Gln Pro Asn Leu
    290                 295                 300
Thr Val His Ile Thr Gly Phe Gly Ala Leu Tyr Tyr Gly Gly Glu Ser
305                 310                 315                 320
Gln Asn Asp Leu Arg Glu Ala Arg Val Lys Ile Ile Ser Asp Asp Val
                325                 330                 335
Cys Lys Gln Pro Gln Val Tyr Gly Asn Asp Ile Lys Pro Gly Met Phe
            340                 345                 350
Cys Ala Gly Tyr Met Glu Gly Ile Tyr Asp Ala Cys Arg Gly Asp Ser
        355                 360                 365
Gly Gly Pro Leu Val Thr Arg Asp Leu Lys Asp Thr Trp Tyr Leu Ile
    370                 375                 380
Gly Ile Val Ser Trp Gly Asp Asn Cys Gly Gln Lys Asp Lys Pro Gly
385                 390                 395                 400
Val Tyr Thr Gln Val Thr Tyr Tyr Arg Asn Trp Ile Ala Ser Lys Thr
                405                 410                 415
Gly Ile
```

<210> 19
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 19
aatgagatac cacgtatctt tcagatccct tg 32

<210> 20
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 20
atgagaagta ccgctctgca gcaagagag 29

<210> 21
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 21
attcgcggcc gcttagatgc ctgtttttga agcaat 36

<210> 22
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 22
gacatacgcc ggattagttt gccagaagcc tct 33